Reveal or Conceal?
Employer Learning in the Labor Market for Computer Scientists

Alice H. Wu*

January 13, 2025

Click here for the latest version.

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 existence of an 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 in the absence of employer learning from scientific publications, 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 assortative matching.

*University of Wisconsin - Madison. I am deeply grateful to Larry Katz, Claudia Goldin, Elie Tamer, and David Card for their invaluable guidance and encouragement. I thank Anna Le Sun for exceptional research assistance, and thank Mandy Pallais, Jesse Shapiro, David Cutler, Louis Kaplow, Francesca Miserocchi, Antonio Coran, Ayushi Narayan, Audrey Tiew, Ariel Gomez, Omeed Maghzian, Jacob Weber, four anonymous scientists from industry, and seminar participants for helpful discussions and comments. Erfan Samaei, Maisam Goreish, Dan Ma, and Rishab Srivastava contributed to an earlier database of patent assignees that helped this project as well. This work was supported by research grants from the Lab for Economic Applications and Policy, and Harvard Stone Program in Wealth Distribution, Inequality & Social Policy. All errors are my own.

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). This missing data challenge also makes it difficult to quantify the impact of employer learning on the efficiency of talent allocation, which is a typical outcome of interest 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 dynamic model of employer learning and sorting. I consider the wage setting and task allocation decisions made by firms in an imperfectly competitive labor market. Firms that vary in productivity allocate workers to innovation task and update their beliefs about the research ability of workers based on their innovation outputs. When part of the innovation record is publicly visible, firms face a trade-off between learning and retention: allocating a worker to more innovation tasks helps an incumbent employer identify high-ability workers faster and improve internal productive efficiency, but it also increases the risk that high-ability workers will be recognized and poached by outside employers. The Markov Perfect Bayesian Nash Equilibrium of this model comprises profit-maximizing wages set by firms conditional on the information they have, taking as given the wages set by their competitors. To the best of my knowledge, this is the first dynamic model that introduces information frictions to a monopsony framework as in Card et al. (2018).

This model generates two key testable predictions on how information revelation changes inter-firm mobility: (1) Workers with newly revealed innovation

[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.

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 employers outside. 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.

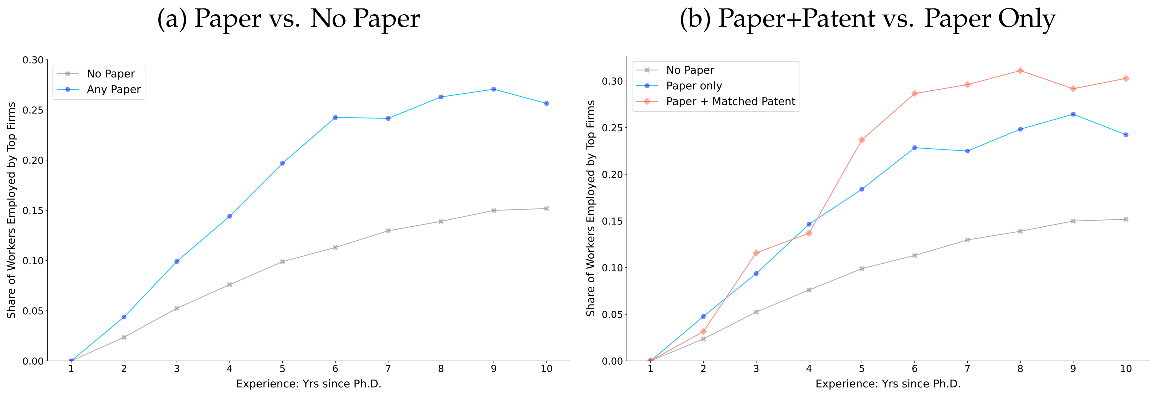
I test for public employer learning by comparing the job mobility of workers who produce a paper with similar coworkers without a paper. Authors of CS papers often get a second chance to move up the job ladder. I measure upward mobility by job movements into top "big tech" firms {Google, Microsoft, IBM, 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 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 higher-ability researchers

[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.

and more productive firms.

Figure 1: Upward Mobility from Nontop to Top Firms



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 feature suggests whether a paper has a matched patent application is revealed later than the paper itself.[7] Workers themselves rarely advertise pending patent applications that have not been published by the patents office.[8]

Figure 1b based on the raw data shows another divergence in upward mobility between workers who produce a paper only or a paper with a matched patent. Comparing similar coworkers at nontop firms, I find that authors of papers with a matched patent are less likely to move than other authors when only the papers are known. But in three years when most patent applications become public information, they are 14% more likely to move to a new firm, and 26% more likely

[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]See Appendix Figure B7. Most workers will sign a non-disclosure agreement, which defines any invention on the job as the employer's proprietary information. Patent applications that have not been published may still be viewed as trade secrets (e.g., *Hyde Corporation v. Huffines* 1958). It is therefore risky for workers to publicly signal patent applications that are still private information of the incumbent employer.

to move to a top firm than their coworkers. This finding is consistent with the model prediction on asymmetric learning: 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.

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 faster positive assortative matching. Workers who produce a paper with a matched patent would experience a 2 pp increase in upward mobility without further ado, and generate a 5-6% increase in innovation production at top firms. However, in the absence of information rent, incumbent firms would assign fewer innovation tasks ex ante, providing a counterforce on the discovery of talent in this counterfactual scenario. This result is similar to the prediction that firms provide less general training when they face higher turnovers (e.g., Acemoglu and Pischke 1998; 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 et al. (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) also 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 worker research ability, denoted by α_i. When new Ph.D.s enter the labor market ($t = 1$), there is public information I_{i1} about person i that is predictive of whether her α_i is high or low, such as education background or earlier publications. Post-Ph.D. employer learning, in contrast, is based on the innovation outputs produced by workers on their jobs and is asymmetric between firms when the incumbent employer has additional information earlier than the outside labor market.

Specifically, the research papers produced by workers become 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. I refer to research papers and their accompanying patent applications as publishable innovation. The likelihood of publishable innovation is determined by a worker's ability and the share of time she can spend on publishable innovation tasks, denoted by τ, as opposed to routine tasks. The task allocation τ is endogenously chosen by her employer.

Firms (j) are heterogeneous in a multidimensional productivity. Some firms are better at attracting clients through publishable innovation or developing them to products, whereas other firms may benefit little from publications and rely more on private research that yields traditional patents unrelated to papers. The production of publishable innovation is supermodular in equilibrium: firms that are more productive in publicly accessible research will set a higher τ, which thus increases the difference in expected output between H and L workers.

Conditional on information about workers, firms make simultaneous offers of a wage $\{w_{itj}\}$ and a time allocation to publishable innovation tasks $\{\tau_{itj}\}$ 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 publishable innovation today, but any successful paper becomes public information and increases the risk of losing the author to other firms 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 publishable 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 nested logit preferences over employers, ϵ_{i1j}, which can be correlated within each nest $G(j) \in$ {Tenure-Track, Postdoc, Top Firms, Nontop Firms} but independent between G's. At $t > 1$, I follow Card et al. (2018) to allow workers to re-enter the labor market and redraw ϵ_{itj} 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 previous 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 wage-posting game between firms. Both the wage w and innovation task τ will increase when a worker publishes a paper or when a patent application from the earlier period becomes public knowledge. The wage increase upon information revelation in equilibrium 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) that there is increased mobility from less productive firms when the labor market receives positive information about workers.

2.2 Model Specification and Equilibrium

I clarify the notations 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 Notations and 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 output while being employed by firm $j(i, t)$.

The output of worker i in period t is summarized by a vector of indicators, $(D_{it}(11), D_{it}(10), D_{it}(00))$, the first two of which represent publishable innovation: $D_{it}(11) = 1$ if i produces any paper with an accompanying patent, whereas $D_{it}(10) = 1$ if i produces papers but none of which are applied as patents. To take into account some firms focus on private research rather than publishable innovation, I introduce $D_{it}(01)$, which indicates if i has any patent application unrelated to papers. Conditional on allocation to publishable innovation tasks τ, high-ability ($\alpha_i = H$) workers are more likely to produce papers than the low-ability ($p_H > p_L$), and are more likely to produce papers with an accompanied patent $\widetilde{p}_H > \widetilde{p}_L$ (Table 1).[9] to be independent from τ. H workers are more likely to patent ($q_H > q_L$), but there are no increasing differences between firms.

[9]When a worker has any paper, $D_{it}(11) + D_{it}(10) = 1$. I require $D_{it}(11) = 1$ and $D_{it}(10) = 0$ when any of her papers has a matched patent application. The probability that a worker of ability α has a matched patent application conditional on having any paper is denoted by $\widetilde{p}_\alpha = E[D_{it}(11)|D_{it}(11) + D_{it}(10) = 1]$. Under this assumption, there are 6 possible innovation outputs:

$$(D_{it}(11), D_{it}(10), D_{it}(01)) \in \{(1, 0, 1), (1, 0, 0), (0, 1, 1), (0, 1, 0), (0, 0, 1), (0, 0, 0)\} \tag{2.1}$$

Table 1: Likelihood of Innovation Output

Innovation	Likelihood | α and innovation task τ
$D_{it}(11)$: Any Paper + Matched Patent	$E[D_{it}(11)\|\alpha, \tau] = p_\alpha \times \widetilde{p}_\alpha \times \tau$
$D_{it}(10)$: Any Paper but no Matched Patent	$E[D_{it}(10)\|\alpha, \tau] = p_\alpha \times (1 - \widetilde{p}_\alpha) \times \tau$
$D_{it}(01)$: Any Patent unrel. to Paper	$E[D_{it}(01)\|\alpha, \tau] = q_\alpha$

Incumbent employer $j(i,t)$ has full access to $(D_{it}(11), D_{it}(10), D_{it}(01))$. However, due to the delayed disclosure of patent applications, the outside labor market initially only observes if there is a research paper, denoted by $D_{it}(11) + D_{it}(10)$, but cannot tell at t if any of the worker's paper is accompanied by a patent application ($D_{it}(11)$ vs. $D_{it}(10)$), or if she has any patent unrelated to papers, $D_{it}(01)$. The information evolution at $t \geq 1$ can be summarized as follows, in which the private $\widetilde{I}_{it}$ becomes public with a one period delay[10]

$$\text{public } I_{i(t+1)} = \underbrace{I_{it} \cup \widetilde{I}_{it}}_{\text{info before } t} \cup \{j(i,t), \underbrace{D_{it}(11) + D_{it}(10)}_{\text{any paper at } t}\} \tag{2.3}$$

$$\text{private } \widetilde{I}_{i(t+1)} = \{(D_{it}(11), D_{it}(10), D_{it}(01))\}$$

Firms are endowed with a baseline productivity $\bar{\phi}_j \in \mathbb{R}^+$, and proportionate returns to each type of innovation, $\left[\phi_j(k)\right]_{k \in \{11,10,01\}}$ with $\phi_j(k) \in \mathbb{R}^+$, all of which are publicly known. Firms that benefit more from publishable innovation have a higher $\phi_j(10)$ or $\phi_j(11)$, while firms that rely more on traditional patenting like Apple have a higher $\phi_j(01)$. Conditional on worker ability α and task allocation τ,

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

$$Pr(I_{i(t+1)}, \widetilde{I}_{i(t+1)} | I_{it}, \widetilde{I}_{it}) = \sum_{\alpha \in \{H,L\}} \underbrace{Pr(\alpha | I_{it}, \widetilde{I}_{it})}_{\text{current belief}} \times \underbrace{Pr(D_{it}(11), D_{it}(10), D_{it}(01) | j(i,t), \alpha)}_{\text{innovation output at } t} \tag{2.2}$$

the expected value from the production at firm j is:[11]

$$Y_j(\alpha,\tau) = \bar{\phi}_j \left(\underbrace{1-\tau}_{\text{routine}} + \underbrace{\sum_{k\in\{11,10,01\}} \phi_j(k) \times E[D_{it}(k)|\alpha,\tau]}_{\text{returns to innovation}} - \underbrace{\zeta(\tau)}_{\text{cost}} \right) \tag{2.4}$$

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 distribution:

$$F(\{\epsilon_{itj}\}) = exp\left(- \sum_{G\in C} \left(\sum_{j\in G} exp(-\rho_G^{-1}\epsilon_{itj}) \right)^{\rho_G} \right) \tag{2.5}$$

where C denotes the set of potential employers a worker can choose from in a given period.[12] The preferences are independent between nests and over time, but can be correlated within a nest G if $\rho_G < 1$. Among the four nests $G(j) \in$ {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:

$$\lambda(I_{it}) = \lambda_{0,G} \times (1 + \lambda_{1,G} \times Pr(H\,|\,I_{it})) \tag{2.6}$$

[11] There is a convex cost of allocating workers to publishable innovation tasks, which may include investment in computing power that often grows in a convex way as employees spend more time on innovation. 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, hire new managers, and establish a new performance evaluation system for workers who are increasingly involved in research.

[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 [54].

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 her employer as follows:

$$j(i,t) = argmax_{j \in C}\ u_{itj} = b \times ln(w_{i\,t\,j}) + \rho_{G(j)} \times \epsilon_{i\,t\,j} \tag{2.7}$$

Assume $b \in (0, \infty)$ and $\forall G : \rho_G \in (0, 1]$ so that the labor market is imperfectly competitive. Their choice probabilities are represented by the well-known nested logit model (McFadden 1973; Imbens and Wooldridge 2007):

$$s_{j\,|\,C} = \underbrace{s_{j\,|\,G(j)}}_{\text{choose } j \in G(j)} \times \underbrace{s_{G(j)|\,C}}_{\text{choose nest } G(j) \in C} \tag{2.8}$$

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:

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

$$\text{Outside } j \neq j(i, t-1) : \ s_j^{(0)}(\{w_{itj'}\}; I_{it}) = \lambda(I_{it}) \times E_C\left[s_{j|C}\right] \tag{2.9}$$

The elasticity of labor supply to a firm is lower among incumbent employees when $\lambda < 1$ (see $\xi_{itj}^{(1)}$ in 7.5). The labor market frictions from λ's provide firms additional monopsony power over incumbent employees relative to new workers.

[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 $z < \overline{z}$, where $\Phi(\overline{z}) = \lambda$. 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).

2.2.3 Employers' Problem

I focus on how employers set wages and allocate workers to innovation tasks in an intermediary period $t \in \{2,,...,T-1\}$. The complete backward induction is presented in the Appendix A1. Employer j's value function is summed over its incumbent employees and potential recruits from other firms (7.13).

For an incumbent employee, employer j solves:

$$v^{(1)}_{tj}(I_{it}, \widetilde{I}_{it}) = max_{\boldsymbol{w},\boldsymbol{\tau}} \underbrace{s^{(1)}_j(\boldsymbol{w}, w_{-j}; I_{it})}_{\text{expected labor supply}} \times \underbrace{\left(E_{\alpha|I_{it}\cup\widetilde{I}_{it}}[Y_j(\alpha, \boldsymbol{\tau})] + \beta\, E[v^{(1)}_{(t+1)j}(I', \widetilde{I}') \mid \boldsymbol{\tau}] - \boldsymbol{w}\right)}_{\text{flow profit \& discounted continuation value, net wage}} \tag{2.10}$$

in which w_{-j} are wages posted simultaneously by other firms, taking as given by j.[15] The continuation value equals the value from an incumbent worker $v^{(1)}_{(t+1)j}$, expected over the innovation outputs that will be produced in the current period (see equation 7.15).[16] 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 $\xi^{(1)}_{itj}$ (7.5):[17]

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

Task allocation is chosen to maximize the expected returns to publishable innovation today and dynamic returns to the continuation value. Employers consider how task allocations would affect public information about a worker and her

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

[16]Information evolves according to the worker's innovation outputs at t as in (2.3), with $I' = I_{it} \cup \widetilde{I}_{it} \cup \{D_{it}(11) + D_{it}(10)\}$, $\widetilde{I}' = \{(D_{it}(k))_{k=11,10,01}\}$.

[17]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.

turnover tomorrow.

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

$$\boldsymbol{\tau}^{*}_{tj}(I, \widetilde{I}) := \underbrace{\frac{1}{\zeta} \times E_{\alpha|I\cup\widetilde{I}}\left[-1 + \sum_{k\in\{11,10,01\}} \phi_j(k) \times \frac{\partial E[D_{it}(k)|\alpha,\tau]}{\partial\tau}\right]}_{\text{return to innovation today}} + \frac{\beta/\bar{\phi}_j}{\zeta} \times \underbrace{\frac{\partial E[v^{(1)}_{(t+1)j}(I', \widetilde{I}')|\tau]}{\partial\tau}}_{\text{dynamic return}} \tag{2.13}$$

The derivative of continuation value over task allocation can be negative if workers who publish a research paper are likely to be poached away (7.19). The lower innovation assignment 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 information I_{it}. Its value function is expected over the unknown private signals $\widetilde{I}_{it}$:

$$v^{(0)}_{tj}(I_{it}) = max_{\boldsymbol{w},\boldsymbol{\tau}} E_{\widetilde{\boldsymbol{I}}|I_{it}}\Big[\underbrace{s^{(0)}_j(\boldsymbol{w}, w_{-j}; I_{it})}_{\text{expected labor supply}} \times \underbrace{\left(E_{\alpha|I_{it}\cup\widetilde{\boldsymbol{I}}}[Y_j(\alpha,\tau)|I_{it}\cup\widetilde{\boldsymbol{I}}] + \beta\, E[v^{(1)}_{(t+1)j}(I', \widetilde{I}')\,|\boldsymbol{\tau}] - \boldsymbol{w}\right)}_{\text{MRPL \& discounted continuation value, net wage}}\Big] \tag{2.14}$$

The wage for a new worker is a weighted average of monopsonistic wages (marked down by elasticity $\xi^{(0)}_{itj}$ (7.10) conditional on information $(I_{it}, \widetilde{I})$:

$$\begin{aligned} \boldsymbol{w}^{(0)}_{itj} = &\left(1 + E_{\widetilde{I}|I_{iT}}\left[\frac{s^{(0)}_j}{E_{\widetilde{I}|I_{iT}}[s^{(0)}_j]} \times \xi^{(0)}_{iTj}(\widetilde{I})\right]\right)^{-1} \\ &\times E_{\widetilde{I}|I_{iT}}\left[\frac{s^{(0)}_j}{E_{\widetilde{I}|I_{iT}}[s^{(0)}_j]} \times \xi^{(0)}_{iTj}(\widetilde{I}) \times \left(E_{\alpha|I_{it}\cup\widetilde{I}}[Y_j(\alpha,\tau)] + \beta E[v^{(1)}_{(t+1)j}(I', \widetilde{I}')|\tau]\right)\right] \end{aligned} \tag{2.15}$$

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.[18]

[18]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.

When incumbent employers set higher wages for workers with more positive $\widetilde{I}$, outside firms would lower the weights on such $\widetilde{I}$, which reflects the firms taking into account the averse selection of movers under asymmetric information (Gibbons and Katz 1991; Greenwald 1986).

The optimal allocation of new workers to innovation tasks is a weighted average of $\tau^*_{tj}(I, \widetilde{I})$, with the same weight on each unknown $\widetilde{I}$ as in (2.15):

$$\boldsymbol{\tau}^{(0)}_{itj} = E_{\widetilde{I}|I_{it}}\left[\frac{s^{(0)}_j}{E_{\widetilde{I}|I_{it}}[s^{(0)}_j]} \times \tau^*_{tj}(I_{it}, \widetilde{I})\right], \ \tau^* \text{ defined in (2.12)} \tag{2.16}$$

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 wages conditional on their current information about a worker, taking as given the wages posted by other firms.

Definition 1 (Markov Perfect Bayesian Equilibrium Under Imperfect Competition) *In this T-period game between firms, the equilibrium in an imperfectly competitive labor market ($\frac{b}{\rho_G} \in (0, \infty)$) comprises:*

- $t = 1$: *wages $\{w_{1j}(I)\}$ at each firm, given every possible initial information I about a worker;*
- $t = 2, \ldots T$: *wages $\{w^{(1)}_{tj}(I, \widetilde{I})\}$ for incumbent employees at each firm given every possible public information I and private information $\widetilde{I}$ at t, and wages $\{w^{(0)}_{tj}(I)\}$ for workers from other firms given only public information;*

that are set by each firm to maximize its own profits, taking as given the wages set by other firms.[19]

Let $\boldsymbol{w}$ denote the vector of wages in Definition 1 set by all firms for all possible information sets, and $\boldsymbol{s} = s(\boldsymbol{w})$ denote the vector of expected labor supply evaluated

[19]Specifically, employers solve (7.20) at $t = 1$, (2.10, 2.14) at $t = 2, \ldots, (T-1)$, and (7.3, 7.8) at $t = T$. The expected labor supply is shaped by workers who choose their employers according to (2.7) conditional on the wages posted by all potential employers.

at equilibrium wages $\boldsymbol{w}$.[20] I can write the equilibrium as a system of equations:

$$\boldsymbol{s} = s(w(\boldsymbol{s})) \tag{2.17}$$

in which the wage function w denotes the optimal wage as in (2.11, 2.15), and s denotes the conditional choice probability in (2.9) given information $(I, \widetilde{I})$.

I show the existence and uniqueness (up to positive scaling of the wages) of the MPBNE in Proposition 1 (Appendix A2). The proof closely follows the discussion of discrete dynamic games with incomplete information in Rust (1994): $s \circ w$ is proved to be a contracting mapping, with a unique fixed point $\boldsymbol{s}$ (2.17) that represents the allocation of workers between firms in equilibrium.

The allocation to innovation 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 (Proposition 2 in Appendix A2), workers who are not credit-constrained bear all costs of innovation tasks and are paid their full marginal product of labor as in Becker (1964).

2.3 Model Predictions

I discuss the implications of information revelation on inter-firm mobility. The predictions are derived from the equilibrium under the following assumptions:

A1: The labor market is imperfectly competitive: $b/\rho_G \in (0, \infty)$.

A2: In the nest G = nontop firms, the probability of re-entering the labor market (2.6) satisfies: $\forall$ information I, I': $Pr(H|I) > Pr(H|I') \rightarrow \lambda(I) > \lambda(I')$.

[20] Given information $(I, \widetilde{I})$ in period t, the equilibrium wage at firm j:

$$\boldsymbol{w}_{tj}(I, \widetilde{I}) = \begin{cases} \boldsymbol{w}_{1j}(I) & \text{if } t = 1\text{, as in equations (7.22)} \\ \boldsymbol{w}_{tj}^{(1)}(I, \widetilde{I}) & \text{if } t > 1 \text{ and } j = j(i, t-1)\text{, as in equations (2.11, 7.6)} \\ \boldsymbol{w}_{tj}^{(0)}(I) & \text{if } t > 1 \text{ and } j \neq j(i, t-1)\text{, as in equations (2.15, 7.11)} \end{cases}$$

Prediction 1 (Job Mobility in Response to Public Information) *Conditional on public information about research ability, workers who produce any public innovation while being employed by less productive firms are*

a) more likely to move to a new employer,

b) more likely to move to an employer with higher innovation productivity

than coworkers without innovation.

Public signals such as a research paper or the delayed revelation of a patent application improve the market belief about if a worker is H-ability. The equilibrium wages across firms increase in response to the positive public signals, but importantly, the wage increase is higher at more productive firms due to the complementarity between firms and workers in innovation. Firms with higher $\phi_j(1\cdot)$ in innovation assign more innovation tasks to the same worker, and set disproportionately higher wages relative to less productive firms. As a result, the model predicts an increase in inter-firm mobility, and an increase in upward mobility for workers who release positive signals at lower-productivity firms.

Prediction 2 (Job Mobility under Asymmetric Information: $D_{it}(11)$ vs. $D_{it}(10)$) *Workers who have produced a high-quality paper with a matched patent, i.e.* $D_{it}(11) = 1$, *while being employed by less productive firms are*

a) less likely to leave their incumbent employers when the presence of a matched patent $D_{it}(11) = 1$ *is private information;*

b) more likely to move and move upward after $D_{it}(11) = 1$ *is revealed.*

than coworkers with papers but no matched patents $D_{it}(10) = 1$.

The second prediction relies on the assumption that at index period t the outside labor market may see a paper but cannot differentiate between $D_{it}(10)$ and $D_{it}(11)$ (see 2.3). Incumbent employers set a higher wage based on a more favorable private belief $Pr(H|I \cup \widetilde{I}) > Pr(H|I)$ when $D_{it}(11) = 1$, and therefore reduces the turnover of $D_{it}(11) = 1$ workers relative to coworkers who produce papers without a matched patent $D_{it}(10) = 1$. Once the matched patent is revealed in the next period, Prediction 1 re-applies.

Prediction 3 (Job Mobility under Asymmetric Information: $D_{it}(01)$) *Workers who produce a patent application unrelated to any paper, $D_{it}(01) = 1$, experience a delayed increase in job mobility relative to similar coworkers with $D_{it}(01) = 0$.*

The outside labor market does not observe $D_{it}(01)$ until the next period (2.3). As a result, workers who produce any patent application unrelated to papers are also likely to move later than their coworkers without a patent application, as in Prediction 2.

I will test the predictions from the equilibrium in Section 4, and estimate the model to quantify the role of employer learning on job mobility and productivity in Section 5.

3 Data

I collected data on the career trajectories and research outputs of Ph.D. computer scientists. This section discusses the data sources, the matching between Ph.D. dissertation records and public LinkedIn profiles, and measures of on-the-job research that includes 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.[21] The share of new CS Ph.D.s enter-

[21]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 directly or as research engineers who also publish

ing the industry as opposed to academia has been increasing over the past 20 years and exceeding 50% since 2017 (see 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.[22] Each dissertation record provides the full name of the doctoral recipient, school, and year of PhD.[23]

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

To gather information on the career progression of CS Ph.D.'s, I develop 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 submit a web query that searches by the person's full name, PhD institution, and degree information.[24] About 51% queries returned at least one LinkedIn profile, and there are about 41,000 fully matched profiles in total.[25]

Each profile is formatted as a résumé. The program collects public information such as profile summary, education background, and employment history. I construct a longitudinal dataset of post-Ph.D. employment history for the fully matched LinkedIn profiles. On average a person has 2.1 industry employers, 0.2 postdoc employers, and 0.3 academic (tenure-track) employers after Ph.D. (Table

papers.

[22]The top schools are identified from the ranking of computer science institutions in the U.S. at CSRankings, which is developed and maintained by Emery Berger at UMass Amherst.

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

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

[25]See Appendix 7 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/`).

B5). For each person×year, I record the primary employer and job title.[26] The person×year panel has about 647,000 observations.

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 productivity of Ph.D. computer scientists, I collect papers that are published in 80 CS conferences and two machine learning journals, which are used to rank CS departments across all areas in CSRankings. I search for papers at each conference/journal×year on Scopus, a large-scale publication database produced by Elsevier.[27] Each paper comes with a complete list of authors and their affiliations, which indicate the employer of an author at the time of publication. 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).

To match papers with individuals' education and employment history, I developed a script to clean and harmonize the names of author affiliations from Scopus, and the names of Ph.D. schools and employers from LinkedIn. A paper

[26] If there are 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 on a job since the earliest date.

[27] I am especially grateful to Anna Le Sun (Berkeley/Stanford) for her help with the data collection via Scopus Search API.

matched to an author's Ph.D. institution by (author name, affiliation, year of publication) 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 incumbent employer at the time of publication.[28] 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 do so (Table B6).

3.3.2 Paper-Patent Matches

Firms often protect inventions that are disclosed in a research paper through patents. 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).[29]
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.[30]

[28]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.

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

[30]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.

For each paper, I sort potential patent matches that satisfy the criterion above by the number of shared team members, the distance between embeddings, and the time difference between the earliest filing of a patent application and the publication date of the paper, in ascending order. I keep the first patent application returned as the best possible match to the paper.

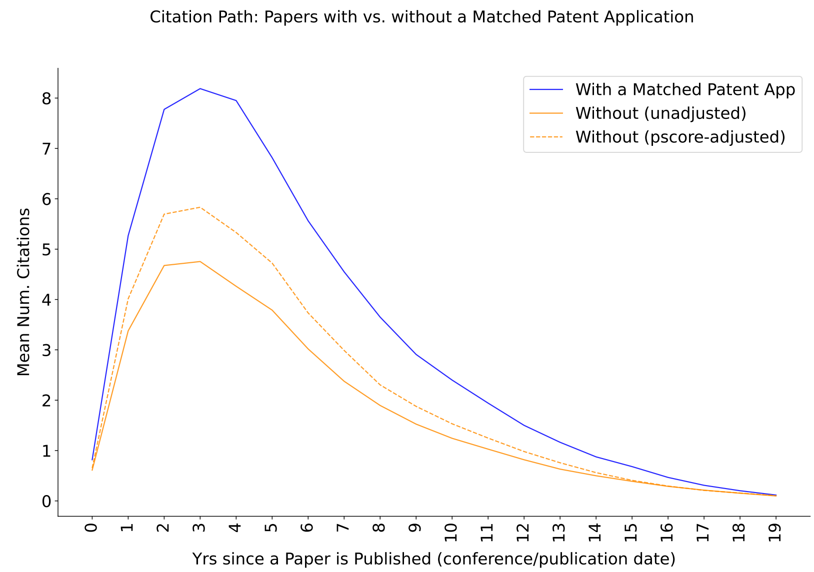
About 25% of papers by matched computer scientists from industry, and 5% of papers by those from academia, are accompanied by a patent application. 90% of the matched patent applications are filed before the research paper shows up at a conference, and the other 10% are filed within 12 months. Table 2 and Appendix Table B7 provides examples of paper-patent matches. They may have different titles, but discuss the same set of research findings. 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 than what one can observe from a paper alone.

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

Papers with a matched patent are higher quality on average. In the first year, they receive roughly the same citations as those without a matched patent. But the gap starts to expand around two years after the paper becomes public, which coincides with the disclosure of patent applications (Appendix Figure B1). The quality difference between papers with and without a matched patent suggests: 1)

Figure 2: Citations Received by 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.

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 the outside market to recognize valuable research and the timing of the divergence in citations is consistent with the revelation of a matched patent application.

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. I require the assignee (firm) on the application to be the same as the inventor's employer as reported 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).

To be consistent with the notation in the model (Table 1), I summarize the innovation output at person-year level by $(D_{it}(11), D_{it}(11), D_{it}(01))$, in which $D_{it}(11) = 1$ if worker i has any paper with a matched patent application in year t, $D_{it}(10) = 1$ if she has paper(s) but none of which is matched to a patent application, and $D_{it}(01) = 1$ if she has any patent application unrelated to CS papers.

4 Empirical Tests for Employer Learning

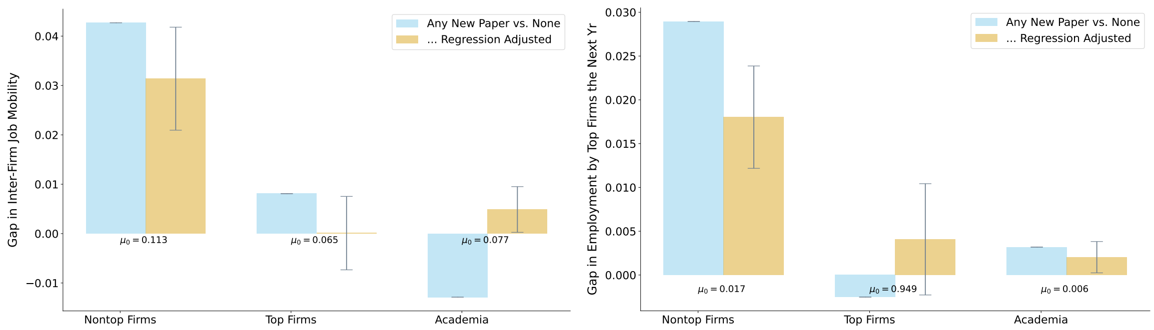
We test the model predictions on how individual job mobility changes with signals about their research ability. There is evidence of public learning (Prediction 1) as inter-firm and upper mobility increases for workers who publish at non-top firms, relative to similar coworkers. To test for asymmetric learning, I leverage the delayed disclosure of patent applications. Initially, authors of papers with a matched patent are less likely to move than authors without a patent application. But once the patent application becomes public, their mobility rates cross over. Authors of papers with a matched patent are also 35% more likely to move to a top tech firm.

4.1 Public Learning: Mobility Responses to CS Papers

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 3(a) first shows the raw differences (blue bars) in inter-firm mobility between these two groups. At nontop firms, workers who produce a paper are on average 4pp 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

Figure 3: 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., firm-year fixed effects and other controls listed under Table 3. μ_0 refers to the mean mobility among workers without a new CS paper.

controls, I also find a significant but smaller increase in mobility when workers in academia publish a new paper, but there is no change in inter-firm mobility for workers who are already employed at the top firms.[31] 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 for workers who publish at nontop firms provides evidence for Prediction 1(a).

Figure 3(b) further shows workers who publish a paper at nontop firms are twice as likely to move to a top firm the next year relative to similar coworkers. 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 noisily estimated.

[31]"Academia" includes postdocs and professors. The raw difference in academia is negative in Figure 3(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 0.5pp significant increase in job mobility in academia.

4.2 Asymmetric Learning: Papers vs. Patent Applications

We consider how job mobility changes with public versus (initially) private signals about research ability:

$$M_{it} = \underbrace{\sum_{k\in\{11,10,01\}} \boldsymbol{\beta_k} \times D_{it}(k)}_{\text{new signals}} + \underbrace{\sum_{k\in\{11,10,01\}} \boldsymbol{\gamma_k} \times \text{Lagged-}D_{it}(k)}_{\text{lagged signals from } [t-3,t-1]} \tag{4.1}$$
$$+ \underbrace{W'_{it}\,\Gamma}_{\text{controls}} + \underbrace{\mu_{j(i,t),\,t}}_{\text{firm-yr}} + \xi_{it}$$

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 (bachelor and Ph.D.), gender (from first names or profile pictures), and time-varying controls such as a polynomial of experience since Ph.D. and position types (e.g., engineers vs. scientists).

The innovation output by each worker i in year t is summarized by $D_{it}(k)$ for $k \in \{11, 10, 01\}$ (Table 1). There are two margins of asymmetric learning. First, given a paper, the outside labor market does not know if it is matched to a patent application that will be disclosed later. That is, $D_{it}(10)$ versus $D_{it}(11)$ cannot be differentiated by outside employers at t. Second, whether a worker has any patent application unrelated to paper, $D_{it}(01)$ would also be private information for the first 18 months.

In three years, 95% of the patent applications will become public information (see Appendix Figure B1). Define Lagged-$D_{it}(11) = 1$ if a worker produces any paper with a matched patent application in the past three years, Lagged-$D_{it}(10) = 1$ if she produces paper(s) during $[t-3, t-1]$ but none of which is matched to a patent, and finally Lagged-$D_{it}(01) = 1$ if she applies for a patent unrelated to CS papers. The lagged indicators for innovation output are public information. We can then translate the model predictions as follows for workers who are employed by less innovative nontop firms:

$$\begin{aligned} &\text{Prediction 1} \longrightarrow \beta_{10} > 0,\ \beta_{11} > 0 \\ &\text{Prediction 2} \longrightarrow \beta_{11} < \beta_{10},\ \text{ whereas } \gamma_{11} > \gamma_{10} \text{ and } \gamma_{11} > 0 \\ &\text{Prediction 3} \longrightarrow \gamma_{01} > 0 \end{aligned} \tag{4.2}$$

in which β_k captures the difference in outcome M_{it} between workers who produce $D_{it}(k) = 1$ and those without neither a paper nor a patent application, and γ_k represents represents the gap between workers who have produced Lagged-$D_{it}(k) = 1$ during $[t-3, t-1]$ and those without any innovation output in the past three years.

We estimate 4.1 separately for workers who are currently employed by nontop firms, top firms, or academia as in Figure 3. For each person, I keep years of full-time employment post Ph.D. and post 2000.[32]

At nontop firms, workers who has a new paper but no matched patent are 3.5 pp ($t \approx 6$) or 26.3% more likely to move than similar coworkers without any innovation (column 1 of Table 3).[33] Workers who produce a paper with a matched patent

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

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

$$E[M_{it}|D_{it}, \text{Lagged-}D_{it}, W_{it}, j(i,t)] = exp\left(\sum_{k\in\{11,10,01\}} \boldsymbol{\beta_k} \times D_{it}(k) + \sum_{k\in\{11,10,01\}} \boldsymbol{\gamma_k} \times \text{Lagged-}D_{it}(k) + W_{it}'\Gamma + \mu_{j(i,t),t}\right) \tag{4.3}$$

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 : $\mathbf{D_{it}(10)}$ vs. $\boldsymbol{D_{it}(11)}$						
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 $\mathbf{[t-3, t-1]}$: Lagged-$\boldsymbol{D_{it}}$(10) vs. Lagged-$\boldsymbol{D_{it}}$(11)						
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						
$D_{it}(01)$	-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-$D_{it}(01)$	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 $j(i,t) \in$ {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.

($D_{it}(11) = 1$) also see a 2.0 pp or 15.0% increase in next-year inter-firm mobility, but to a lesser extent than workers with $D_{it}(10) = 1$. The positive effects of having any paper for next-year mobility among employees at non-top firms is consistent with Figure 3 and provides evidence for Prediction 1 (4.2) on public learning. We find the estimated $\beta_{11} < \beta_{10}$, which suggests nontop firms as incumbent employers can make workers with paper+matched patent stay longer (Prediction 2a, see 4.2). But $(\hat{\beta}_{11} - \hat{\beta}_{10})$ is not significantly different from 0.

There is stronger evidence for asymmetric learning as in Prediction 2b when I examine the relationship between lagged innovation outputs with job mobility. Conditional on the latest innovation, having any paper but no matched patent in the past three years no longer matters for inter-firm mobility even for workers at nontop firms. In contrast, the presence of a paper with a matched patent during $[t-3, t-1]$, i.e. Lagged-$D_{it}(11) = 1$, predicts a $\hat{\gamma}_{11} = 2.0$-pp 0r 14% significant increase in job movement at t. Since Lagged-$D_{it}(10)$ vs. Lagged-$D_{it}(11)$ represent signals that are revealed to the public with a delay, the positive estimate for γ_{11} and the finding that $\gamma_{11} > \gamma_{10}$ supports Prediction 2b (see 4.2).

Papers in index year t or the past three years do not predict a job movement out of a top firm, which is 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 relative to coworkers without a paper. Whether a paper has a matched patent or not does not matter in academia, where less than 5% of CS papers are filed as patent applications.[34]

Columns 4-6 of Table 3 presents the estimates of (4.1) for $M_{it} = 1[j(i, t+1) \in \text{Top}]$. For workers at nontop firms, this outcome represents upward mobility to a

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

top firm in the industry.[35] We 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 1pp or ?% 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 represents a wage increase that I show formally in Appendix Table ?. We do not find evidence that CS papers (and matched patents) matter for retention or movement between top firms (column 5). It is consistent with the model predictions that having a CS paper and a higher-quality paper with a matched patent matter more for workers who are outside the top firms.[36]

Finally, I show a delayed mobility response to patent applications that are unrelated to CS papers, as in Prediction 3. Workers who file new patent applications are less likely to leave their incumbent employers in the same year ($\hat{\beta}_{01} < 0$ 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), supporting (4.2) implied by Prediction 3. In comparison with the estimates on CS papers, traditional patent applications are less predictive of job movements for computer scientists. This feature is not surprising given the emphasis of publication record in recruiting of 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

[35]This definition of upward mobility is imperfect. There may be smaller, more innovative firms than the tech giants. We show results on alternative upward mobility outcomes in Appendix Table ?.

[36]Equilibrium wages are increasing in a firm's (innovation) productivity. Workers who are revealed to be good researchers are more easily lured away by more productive employers that can offer a higher wage.

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. The mobility responses to signals from CS papers are stronger for less experienced individuals (Appendix Figure C1). Alternative tests that exploit within-person variation in innovation production also provide similar evidence of employer learning, which I discuss in Appendix C (Table C2).

4.3 Additional Evidence of Learning: Wage Growth and Promotions

The evidence of symmetric and asymmetric learning is not limited to the inter-firm mobility outcomes I present above. An alternative definition of upward mobility is moving to a higher-paid firm. 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. We consider a change in job titles as a promotion if the new title includes keywords such as "senior".[37] We 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 $D_{it}(10)$ and $D_{it}(11)$ without any delay, I do not find that workers who 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 $D_{it}(11)$ workers at nontop firms are 1 pp more likely than $D_{it}(10) = 1$ workers to become a research scientist the next year.[38] 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).

In summary, I present empirical evidence of both symmetric and asymmetric employer learning by estimating the changes in job mobility upon information revelation. Our main results support the model-based predictions on inter-firm mobility among productive employees who are not at the top of the job ladder.

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

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

5 Quantitative Analysis

How much does employer learning matter for reducing misallocation of talent? To answer this question, I estimate a structural version of the model of employer learning and sorting formulated in Section 2. Employer learning from on-the-job research matters as much as learning from initial information such as PhD ranking. Without employers' belief update from public research records, innovation output by computer scientists would drop by 16%. Disclosing patent applications one year faster has a small but positive impact on overall innovation, which is driven by faster sorting of H-ability workers into top firms.

5.1 Model Estimation

I discuss the structural parameters and present the estimation procedure that is based on the nested fixed point algorithm (Rust 1987; Rust 1994). The goal is to find estimates that maximize the joint likelihood of job histories and innovation outputs in the first five years of their post-PhD career.[39] I show the model fit and the learning process evaluated at the maximum likelihood estimates.

5.1.1 Parameters and Identification

There are four sets of model parameters that govern each of the following in the structural model: (1) common prior conditional on initial information, (2) labor supply, (3) firm productivity, and (4) worker productivity.

First, I assume the labor market forms a common prior based on initial information I_{i1} about a new Ph.D., comprising {PhD School Rank, Num. Papers before

[39]The choice of $T = 5$ allows me to build a balanced panel for $18,860$ The first few years are particularly important for employer learning, as evidenced by Altonji and Pierret (2001) and Farber and Gibbons (1996)) and also verified in this labor market in Figure C1.

Table 4: Overview of Model Parameters

Parameters	Description
I. Common Prior	
δ	Given initial information I_{i1}, prior: $$Pr(\alpha_i = H \| I_{i1};\, \delta) = \frac{exp\,(\delta' X(I_{i1}))}{1 + exp\,(\delta' X(I_{i1}))} \quad (5.1)$$
II. Labor Supply	
$b, \{\rho_G\}, \{\eta_{\cdot G}\}$	Worker's utility (2.7): weight on log wage and GEV-preferences (2.5)
$\{\lambda_{\cdot,G}\}, \{\Lambda_{\cdot}\}$	Prob. re-entering the labor market (2.6) and moving bet. academia and industry.
III. Firm Productivity	
$\overline{\phi}_j$	Baseline productivity of 16 groups of employers (Appendix D)
$\{\phi_j(k) : k = 10, 01\}$	Returns to each type of innovation: patent $\phi_j(01)$ calibrated, $\phi_j(11)$ is assumed to be a weighted avg of $\phi_j(10)$ and $\phi_j(01)$
IV. Worker Productivity	
$p_\alpha, \widetilde{p}_\alpha, q_\alpha$	Ability-specific productivity in innovation (Table 1)

PhD, Nest of the First Job}:[40] I observe each person's education and publication history before PhD. The initial placement $G_{i1} \in$ {Tenure-track, Postdoc, Top Firms, Nontop Firms} is also part of I_{i1} to allow employers have additional information that matters for the initial sorting between academia and industry.[41] The identification of δ in (5.1) relies on within-firm variation in initial background between coworkers.

The second set of parameters in Table 4 enters a nested logit model for workers' choices between differentiated employers (2.8). The ratio $\frac{b}{\rho_G}$ governs the elasticity of labor supply (7.5) for employees in nest G. In addition, the parameters $\{\lambda_{\cdot,G}\}$ decide the rate at which workers can get on the labor market and search for jobs at $t > 1$, and $\{\Lambda_{\cdot}\}$ decide if academic employers are open to workers from industry

[40]Let r_i denote the rank of PhD school, n_i denote the number of her papers before PhD, and $G_{i1} \in \{0, 1, 2, 3\}$ denote the nest of her employer at $t = 1$, which correspond to {Tenure-track, Postdoc, Top Firms, Nontop Firms}. I define a vector of controls: $X(I_{i1}) = \left(r_i, r_i^2, n_i, n_i^2, 1[G_{i1} = 0], 1[G_{i1} = 1], 1[G_{i1} = 2], 1[G_{i1} = 3]\right)$.

[41]The underlying assumption is that any information observed by employers but not us can be absorbed by the nest of a person's first job.

and vice versa. Parameters on labor supply 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 j's and allow heterogeneous productivity between j but not within[42]. As shown in Appendix Table D2, the labor market has a nested structure. There are two j's (henceforth employer) on the tenure-track, two postdoc employers, six top firms (each as its own j), and six nontop employers that are grouped based on their patenting activity.[43] The baseline productivity $\overline{\phi}_j$ matters for the average wage level and thus the size of j. Returns to CS papers, denoted by $\phi_j(10)$, matter for the allocation to paper-related innovation tasks, and are identified from movers who become more (less) likely to publish when moving to a higher (lower) $\phi_j(10)$ employer.

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

5.1.2 Estimation Procedure

Let δ denote the parameters in common prior (5.1), and Γ denote the rest of the structural parameters to be estimated in Table 4. Given data on the each person's job history $\{j(i,t)\}$ and innovation outputs $d_{it} := [d_{it}(11), d_{it}(10), d_{it}(01)]$, I search

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

[43]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 patent productivity $\phi_j(01)$ (see Table D2).

for estimates of (δ, Γ) that solve:

$$max_{(\boldsymbol{\delta},\boldsymbol{\Gamma})}\, ln\left(\prod_i L_i(\{j(i,t),d_{it}\}\,|I_{i1};\delta,\Gamma)\right) \tag{5.2}$$
$$= \sum_i ln\left(\sum_{\boldsymbol{\alpha}} \underbrace{Pr(\alpha|I_{i1};\delta)}_{\text{prior}} \times L_i(\{j(i,t),d_{it}\}\,|I_{i1},\boldsymbol{\alpha};\Gamma)\right)$$
$$\text{in which } L_i(\cdot|I_{i1},\boldsymbol{\alpha};\Gamma) = \prod_t \underbrace{s_{itj(i,t)}(I_{it},\widetilde{I}_{it};\Gamma)}_{\text{labor supply}} \times \underbrace{Pr(D_{it}=d_{it}|\boldsymbol{\alpha},\tau_{itj(i,t)};\Gamma)}_{\text{innovation output}}$$

in which information evolves according to (2.3), and the unobserved ability α_i is treated as a random effect. Labor supply conditional on information is evaluated at the MPBNE in Definition 1, which is solved as the fixed point given a guess for Γ. $\{\tau_{itj(i,t)}\}$ are the optimal task allocations set by employers at the equilibrium.[44]

Following Rust (1987), our nested fixed point algorithm contains 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 model parameters 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 have reached the fixed point $s_{tj}(I, \widetilde{I}; \Gamma)$:[45]

$$s_{tj}(I,\widetilde{I};\Gamma) = s\left(w\left(s_{tj}(I,\widetilde{I};\Gamma)\right)\right)$$

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

[44]See equations 2.12, **??** for the optimal task allocations chosen by firms at $t = 2, 3, 4$.

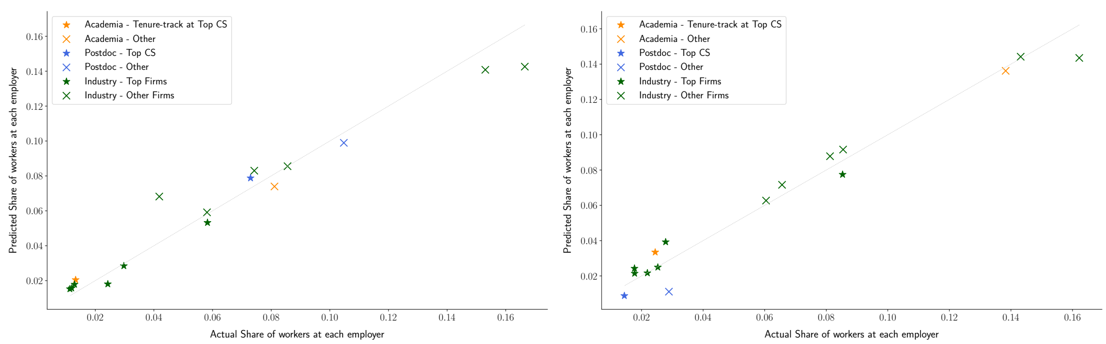
[45]See Proposition 1 for the existence of the fixed point (2.17).

5.1.3 Estimation Results

I estimate the structural parameters on a balanced, five-year panel of 18,860 workers who obtained a PhD between 2005 and 2018. This sample is comparable to the full sample that I use to test for employer learning in Section 5 (see Table D3). Table D1 presents $(\hat{\delta}^{MLE}, \hat{\Gamma}^{MLE})$, the maximum-likelihood estimates of structural parameters. The predicted share of workers at each employer, found as the fixed point (2.17) given $\hat{\Gamma}^{MLE}$, falls roughly on the 45-degree line that matches with the actual shares, at different periods shown in Figure 4.

Figure 4: Model Fit: Allocation of Workers across Employers, $\hat{s}_{tj}$ vs. s_{tj}

(a) $t = 1$ (b) $t = 5$

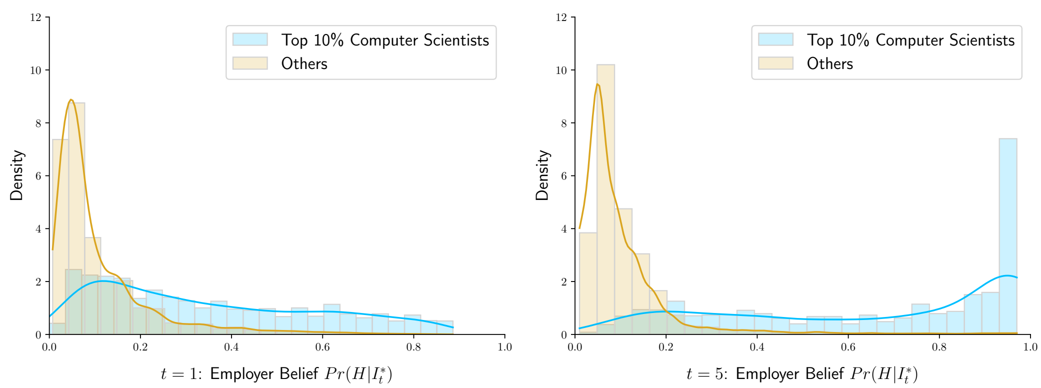


Notes: This figure shows the predicted share of workers at each employer (group) $\hat{s}_{tj}$ against the actual share s_{tj}, at $t = 1$ and $t = 5$. Given the estimated parameters in Table D1, I forward simulate the employment path and innovation outputs by each worker in the balanced sample, holding 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 producing 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 much smaller than in papers.[46] Employers Bayesian update beliefs about a worker based on the innovation outputs they observe. I rank computer scientists by their cumulative citations and total number of papers and patent applications five years after PhD. Figure 5 displays the distribution of employer beliefs, separately for the top 10% computer scientists (as a proxy for H) versus the bottom 90%. At $t = 1$ beliefs about these two groups overlap substantially, suggesting many workers who will be in the top 10% look similar to others initially. But employers appear to tell them apart quickly based on the research outputs they produce. At $t = 5$, there is a more ovbious divergence of beliefs about the bottom 90% versus the top 10%.

Figure 5: Distribution of Employer Beliefs: Top 10% Computer Scientists versus Others

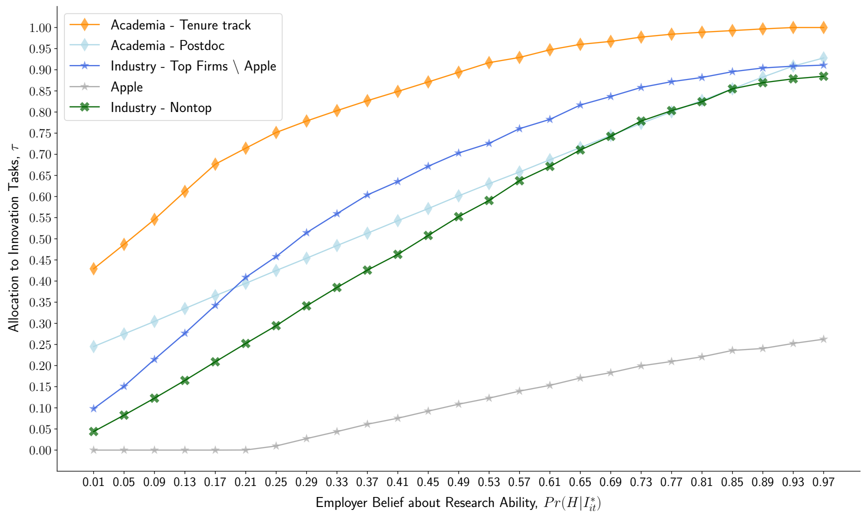
(a) $t = 1$ (b) $t = 5$



H-ability workers produce more papers at tenure-track employers, which have the highest returns to papers ($\phi_j(10)$ in Table D2) and assign more innovation tasks given any employer belief (Figure 6). 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

[46] 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.

Figure 6: Allocation to Innovation Task against Employer Belief



chance of having H-ability but are not fully discovered yet. These potential H-ability 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 talent 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, $D_{it}(01)$;
2. Employer learning from papers with matched patents, $D_{it}(11)$;
3. Employer learning from research papers, $D_{it}(10) + D_{it}(11)$;
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 influence the allocation of talent by changing the evolution of employer beliefs.[47]

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 H vs. L between firms. Figure 7(a) shows how this outcome would change when I shut down the mechanisms one by one (in the order above), relative to the benchmark where all mechanisms are at play.[48] 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. The first substantial drop in publication rate occurs when I shut down employer learning from CS papers. That is, employers no longer update their beliefs based on whether workers have produced a paper. As a result, employers do not assign more innovation tasks to workers who publish, nor are H-ability sorted into more productive firms as efficiently as before. Together, employer learning from innovation outputs $\{(D_{it}(11), D_{it}(10), D_{it}(01)\}$ accounts for the 15.8% of the overall publication rate.[49] Figure D1 further shows that top firms and academic employers experience a larger loss in innovation when they do not learn from the innovation output of the workers.

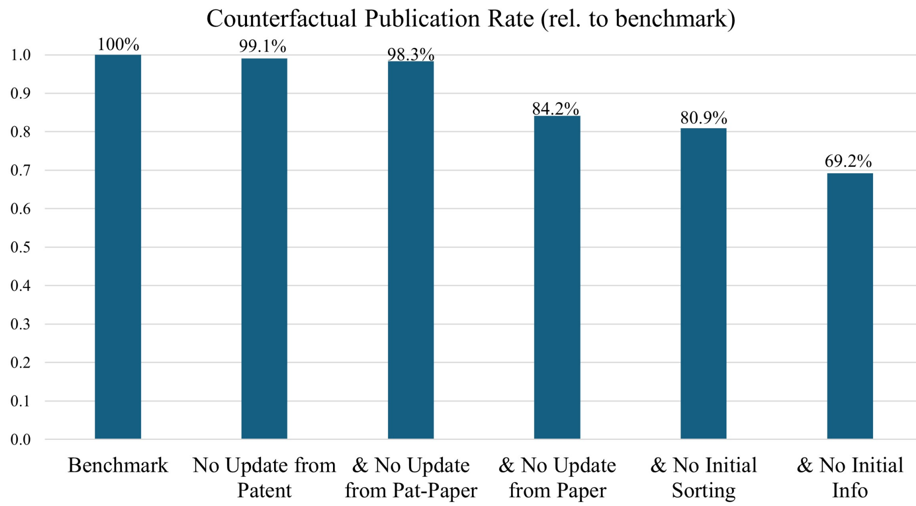
Shutting down initial sorting between nests further reduces the publication rate by 3.3%, while other initial information such as PhD school and papers before

[47]For example, if employers do not update their beliefs based on papers produced by workers, they would not assign additional innovation tasks internally to employees who publish, and authors of papers would also not receive higher wage offers from other firms than coworkers without a paper.

[48]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.

[49]Appendix D3 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.

Figure 7: Decomposition of Publication Rate (Efficiency of Talent Allocation)



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 worker heterogeneity.[50]

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).[51]

I compute the counterfactual publication outcome under 2^5 possible sets of

[50] 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.

[51] 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 et al. (2021) also uses Shapley value to decompose the variance of earning inequality on multiple sources of variation in their counterfactual analysis.

mechanisms.[52] The Shapley value of mechanism $m \in \mathbb{M} = \{1, 2, 3, 4, 5\}$ is:

$$SV_m = \sum_{S \subseteq \mathbb{M} \backslash \{m\}} \frac{|S|! \times (|M| - |S| - 1)!}{|M|!} \times \underbrace{\left(E\left[p_{it} | S \cup \{m\}; \hat{\delta}, \hat{\Gamma}\right] - E\left[p_{it} | S; ; \hat{\delta}, \hat{\Gamma}\right] \right)}_{\text{Change in publication rate when adding mechanism } m} \tag{5.3}$$

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%. Initial information ranks second with an explanatory power of 40.2%. Initial sorting based on information seen by employers but not us matters much less with a value of 2%. Learning from patent applications which are not disclosed immediately to outside employers explain the remaining 8%, which are smaller than learning from papers 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.

[52]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.

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 innovation 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 $D_{it}(11)$, $D_{it}(10)$, and $D_{it}(01)$ are disclosed simultaneously. That is, incumbent employers no longer hold private information for one period. 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 with the same prior $Pr(H|I_{i1})$, and find the equilibrium wages and task allocations at each employer under simultaneous information disclosure.

Figure 8: Upward Mobility from Non-top to Top Firms, Asymmetric vs. Symmetric (dashed)

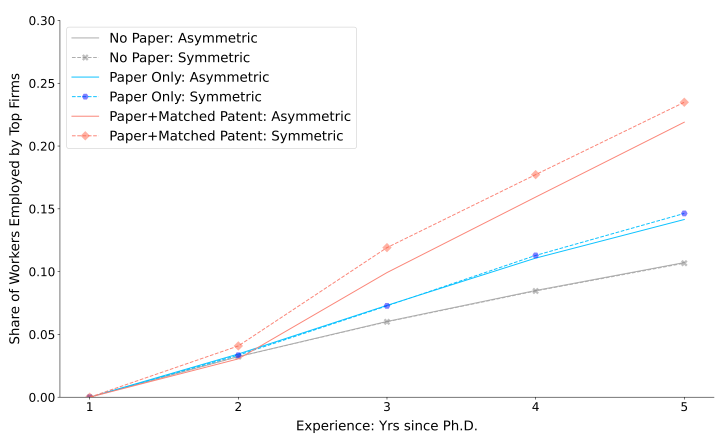


Figure 8 displays the upward mobility for workers who start at nontop firms but produce different innovation output, under the asymmetric benchmark versus the symmetric counterfactual. Workers with output $D_{it}(11) = 1$ can be told apart immediately from workers with a paper only $D_{it}(10) = 1$ under the counterfactual. Relative to the asymmetric benchmark, they move to top firms more quickly. In contrast, workers who only have a paper, $D_{it}(10) = 1$, or no paper at all are as likely to move upward as before.[53]

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%

The overall publication rate would be 1% higher under the symmetric disclosure (Table 6). Top firms benefit from faster information disclosure, experiencing 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 innovation tasks. To decompose the change, I hold fixed the task allocation decision made by firms under asymmetric learning in the simulations for the symmetric counterfactual. The last set of results in Table 6 shows that CS papers would

[53] $D_{it}(10) = 1$ 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.

increase by 1.3% rather than 0.97% when employers do not adjust their task allocations (τ) in response to the reduction of asymmetric information. In other words, faster positive assortative matching 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 tests for employer learning about worker ability and quantifies the role of learning in improving the allocation of talent in the labor market for computer scientists. I build a new dataset that combines the employment histories of newly minted Ph.D.'s in computer science with information on their publications in major conference proceedings and their patent applications. The matched data allows me to 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. To test for asymmetric learning, I exploit a patent law that delays the disclosure of patent applications. Higher-quality papers often coincide with a closely related patent application, but the fact of filing remains private for 18 months. Authors of such papers experience a delayed increase in inter-firm and upward mobility. Conditional on origin firm and observable characteristics, they are less likely to leave the incumbent firms with private information immediately, but once the patent applications become public, they experience a strong increase in inter-firm and upward mobility from nontop firms to top firms in the industry.

The mobility changes around the publication of a CS paper or patent application are consistent with predictions from the dynamic framework of employer learning and sorting in this paper, which introduces information frictions about talent into an imperfectly competitive labor market. 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 faster 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).

This paper suggests that even for a high-skilled group with strong credentials, information frictions are prevalent and result in substantial misallocation of workers between and within firms. The framework of employer learning under imperfect competition may help analyze in the role of information frictions in other 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.

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.

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 ($D_{it}(10) + D_{it}(11) \in \{0, 1\}$) is known by the beginning of $t = 2$; whether a paper from $t = 1$ has a matched patent application ($D_{it}(11)$ vs. $D_{it}(10)$), and whether there is a patent unrelated to paper ($D_{it}(01)$), 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 (2.7) at $t = 1$.
   (c) Innovation outputs (**??**)

   $$(D_{it}(11), D_{it}(10), D_{it}(01)) \in \{(1, 0, 1), (1, 0, 0), (0, 1, 1), (0, 1, 0), (0, 0, 1), (0, 0, 0)\}$$

   are realized by the end of $t = 1$ and are fully known to i's incumbent employer $j(i, 1)$.
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 (2.14).
   (b) Workers re-enter the labor market with probability (2.6). 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 (2.7) 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 begin-

ning of $t = 3$ evolve according to (2.3):

$$\text{Public } I_{i3} = I_{i2} \cup \widetilde{I}_{i2} \cup \{D_{i2}(11) + D_{i2}(10)\}$$
$$\text{Private } \widetilde{I}_{i3} = \{(D_{i2}(11), D_{i2}(10), D_{i2}(01))\}$$

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 2.7). 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 (2.5).[54]

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:

$$s_{j\,|\,C} = \underbrace{s_{j\,|\,G(j)}}_{\text{choose } j\in G(j)} \times \underbrace{s_{G(j)|\,C}}_{\text{choose nest } G(j)\in C} \tag{7.1}$$
$$\forall G: \; s_{G|C} = 1[G \in C] \times \frac{exp(\eta_G(\pi) + \rho_G \times W_G)}{\sum_{G'\in C} exp(\eta_{G'}(\pi) + \rho_{G'} \times W_{G'})}$$
$$\forall j \in G: \; s_{j|G} = \frac{exp(\frac{b}{\rho_G} ln(w_j))}{exp(W_G)}$$

[54] 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.

Backward Induction:

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

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:

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

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

where w_{-j} are wages posted by other employers given public information I_{iT}, taken as given by the j.[55] Public information I_{iT} matters for the probability at which the worker re-enters the labor market (2.6). Take derivatives of the objective function (7.3) over wage $\boldsymbol{w}$:

$$\frac{\partial s_j^{(1)}(\boldsymbol{w}, w_{-j}; I_{iT})}{\partial w} \times \Big(E_{\alpha|I_{iT}\cup\widetilde{I}_{iT}}[Y_j(\alpha, \boldsymbol{\tau})] - w\Big) - s_j^{(1)}(\boldsymbol{w}, w_{-j}; I_{iT}) = 0 \tag{7.4}$$

[55] 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}.

letting $G = G(j)$,

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

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

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

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

$$\xi_{iTj}^{(1)} := \frac{\partial ln(s_j^{(1)}(\boldsymbol{w}, w_{-j}; I_{iT}))}{\partial ln(w)} = \frac{b}{\rho_G} \times E_C[\underbrace{\frac{\lambda_G \times s_{j|G} \times s_{G|C}}{s_j^{(1)}}}_{(c)} \times \underbrace{(1 - \rho_G\, s_{j|G}\, s_{G|C} - (1 - \rho_G) s_{j|G})}_{(d)}] \tag{7.5}$$

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 (7.4), the optimal wage for an incumbent worker is:

$$w_{iTj}^{(1)} = \boldsymbol{w}_{Tj}^{(1)}(w_{-j}; I_{iT}, \widetilde{I}_{iT}) = E_{\alpha|I_{iT} \cup \widetilde{I}_{iT}}[Y_j(\alpha, \boldsymbol{\tau}_{iTj}^{(1)})] \times \underbrace{\xi_{iTj}^{(1)} \times \left(1 + \xi_{iTj}^{(1)}\right)^{-1}}_{\text{markdown}} \tag{7.6}$$

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 innovation tasks, τ,

$$\underbrace{\frac{\partial s_j^{(1)}}{\partial \tau}}_{=0} + s_j^{(1)} \times \frac{\partial E_{\alpha|I_{iT}\cup\widetilde{I}_{iT}}[Y_j(\alpha,\tau)]}{\partial \tau} \geq 0 \tag{7.7}$$

$$\text{define } \tau^*_{Tj}(I,\widetilde{I}) := \frac{1}{\zeta} E_{\alpha|I\cup\widetilde{I}}\left[p_\alpha \widetilde{p}_\alpha \phi_j(11) + p_\alpha(1-\widetilde{p}_\alpha)\phi_j(10) - 1\right]$$
$$\longrightarrow \boldsymbol{\tau}^{(1)}_{iTj} = max\{0, min\{1, \tau^*_{Tj}(I_{iT},\widetilde{I}_{iT})\}\}$$

Outside Workers

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:

$$v^{(0)}_{Tj}(I_{iT}) = max_{\boldsymbol{w},\boldsymbol{\tau}} E_{\widetilde{I}|I_{iT}}\,[\underbrace{s_j^{(0)}(w,w_{-j};I_{iT},\widetilde{I})}_{\text{labor supply}} \times \underbrace{\left(E_{\alpha|I_{iT}\cup\widetilde{I}}[Y_j(\alpha,\boldsymbol{\tau})] - \boldsymbol{w}\right)}_{\text{MRPL net wage}}] \tag{7.8}$$
$$\text{where } s_j^{(0)}(w,w_{-j};I_{iT},\widetilde{I}) = \lambda(I_{iT}) \times E_C\left[s_{j|C}(\boldsymbol{w},w_{(-j)};I_{iT},\widetilde{I})\right]$$

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 (7.8) varies by private information $\widetilde{I}$.

Taking the derivative of (7.8) over the wage $\boldsymbol{w}$ posted by j:

$$E_{\widetilde{I}|I_{iT}}\left[\frac{\partial s_j^{(0)}}{\partial w} \times \left(E_{\alpha|I_{iT}\cup\widetilde{I}}[Y_j(\alpha,\tau_j^{(0)})] - w\right) - s_j^{(0)}(\boldsymbol{w},w_{-j};I_{iT},\widetilde{I})\right] = 0 \tag{7.9}$$

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

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

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

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

$$\xi^{(0)}_{iTj}(\widetilde{I}) := \frac{\partial ln(s^{(0)}_j(\boldsymbol{w}, w_{-j}; I_{iT}, \widetilde{I}))}{\partial ln(w)} = \frac{b}{\rho_G} \times E_C\left[\frac{s_{G|C}}{E_C[s_{G|C}]} \times \left(1 - \rho_G\, s_{j|G}\, s_{G|C} - (1-\rho_G)s_{j|G}\right)\right] \tag{7.10}$$

where $\widetilde{I}$ matters for the wages set by the incumbent employer of this outside worker and thus each choice probability in (7.10). In contrast with the elasticity $\xi^{(1)}_{iTj}$ of an incumbent worker (7.5), $\lambda(I_{iT})$, the probability getting on the market, no longer matters for the elasticity to a new employer j.[56] 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 (7.9), the optimal wage for outside employee i can be written as:[57]

$$\begin{aligned} w^{(0)}_{iTj} = &\left(1 + E_{\widetilde{I}|I_{iT}}\left[\frac{s^{(0)}_j}{E_{\widetilde{I}|I_{iT}}[s^{(0)}_j]} \times \xi^{(0)}_{iTj}(\widetilde{I})\right]\right)^{-1} \\ &\times E_{\widetilde{I}|I_{iT}}\left[\frac{s^{(0)}_j}{E_{\widetilde{I}|I_{iT}}[s^{(0)}_j]} \times \xi^{(0)}_{iTj}(\widetilde{I}) \times E_{\alpha|I_{iT}\cup\widetilde{I}}[Y_j(\alpha,\tau)]\right] \end{aligned} \tag{7.11}$$

Taking the derivative of (7.8) over task allocation τ,

$$\begin{aligned} &\frac{\partial E_{\widetilde{I}|I_{iT}}[s^{(0)}_j \times E_{\alpha|I_{iT}\cup\widetilde{I}}[Y_j(\alpha,\tau)]]}{\partial \tau} \geq 0 \qquad (7.12)\\ \rightarrow &E_{\widetilde{I}|I_{iT}}[s^{(0)}_j \times (-1 + E_{\alpha|I\cup\widetilde{I}}[\phi_j(11)p_\alpha \widetilde{p}_\alpha + \phi_j(10)p_\alpha(1-\widetilde{p}_\alpha)] - \zeta\tau)] \geq 0 \\ \rightarrow &\boldsymbol{\tau}^{(0)}_{iTj} = E_{\widetilde{I}|I_{iT}}\left[\frac{s^{(0)}_j}{E_{\widetilde{I}|I_{iT}}[s^{(0)}_j]} \times \tau^*_{Tj}(I_{iT}, \widetilde{I})\right] \end{aligned}$$

[56] 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.

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

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.

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

$$V_{tj}\left(\bigcup_{\text{worker } i} I_{itj}\right) = \underbrace{\sum_{i:\, j(i,t-1)=j} v^{(1)}_{tj}(I_{it} \cup \widetilde{I}_{it})}_{\text{Incumbent}} + \underbrace{\sum_{i:\, j(i,t-1)\neq j} v^{(0)}_{tj}(I_{it})}_{\text{Workers Outside}} \tag{7.13}$$

Employer j solves the following for incumbent workers:

$$v^{(1)}_{tj}(I_{it}, \widetilde{I}_{it}) = max_{\boldsymbol{w},\boldsymbol{\tau}}\ \underbrace{s^{(1)}_j(\boldsymbol{w},\, w_{-j};\, I_{it})}_{\text{expected labor supply}} \times \underbrace{\left(E_{\alpha|I_{it}\cup\widetilde{I}_{it}}[Y_j(\alpha,\, \boldsymbol{\tau})] + \beta\, E_{\boldsymbol{D}}[v^{(1)}_{(t+1)j}(I', \widetilde{I}') \,|\, \boldsymbol{\tau}] - \boldsymbol{w}\right)}_{\text{MRPL at t \& discounted continuation value, net wage}} \tag{7.14}$$

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

$$\begin{aligned} E_{\boldsymbol{D}}[v^{(1)}_{(t+1)j}(I', \widetilde{I}') \,|\, \boldsymbol{\tau}] &= \sum_{\boldsymbol{D}} Pr(\boldsymbol{D}|I_{it} \cup \widetilde{I}_{it}, \tau) \times v^{(1)}_{(t+1)j}(I'(\boldsymbol{D}), \widetilde{I}'(\boldsymbol{D})) \\ \text{in which } \boldsymbol{D} &= (D_{it}(11), D_{it}(10), D_{it}(01)) \\ Pr(\boldsymbol{D}|I_{it} \cup \widetilde{I}_{it}, \tau) &= \sum_{\alpha} Pr(\alpha|I_{it} \cup \widetilde{I}_{it}) \times Pr(\boldsymbol{D}|\alpha, \tau) \text{ as in Table 1} \\ I'(\boldsymbol{D}) &= I_{it} \cup \widetilde{I}_{it} \cup \{D_{it}(11) + D_{it}(10)\} \\ \widetilde{I}'(\boldsymbol{D}) &= \{\boldsymbol{D}\} \end{aligned} \tag{7.15}$$

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

$$\boldsymbol{w}^{(1)}_{itj} = \left(E_{\alpha|I_{it}\cup\widetilde{I}_{it}}[Y_j(\alpha, \tau)] + \beta\, E[v^{(1)}_{(t+1)j}(I', \widetilde{I}') \,|\, \boldsymbol{\tau}^{(1)}_{itj}]\right) \times \underbrace{\xi^{(1)}_{itj} \times \left(1 + \xi^{(1)}_{itj}\right)^{-1}}_{\text{markdown}} \tag{7.16}$$

The firm-specific labor supply elasticity of an incumbent worker or a new worker can be written the same as equations (7.5) (7.10), 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:

$$\boldsymbol{\tau}^{(1)}_{itj} = max\{0, min\{1, \boldsymbol{\tau}^*_{tj}(I_{it}, \widetilde{I}_{it})\}\} \tag{7.17}$$

$$\boldsymbol{\tau}^*_{tj}(I, \widetilde{I}) := \frac{1}{\zeta} \times \underbrace{E_{\alpha|I\cup\widetilde{I}}\left[-1 + \sum_{k\in\{11,10,01\}} \phi_j(k) \times \frac{\partial E[D_{it}(k)|\alpha, \tau]}{\partial \tau}\right]}_{\text{return to innovation today}} + \frac{\beta/\bar{\phi}_j}{\zeta} \times \underbrace{\frac{\partial E[v^{(1)}_{(t+1)j}(I', \widetilde{I}')|\tau]}{\partial \tau}}_{\text{change in value from stayer}} \tag{7.18}$$

in which the dynamic return to assigning more innovation task today:

$$\begin{aligned}
&\frac{\partial E[v^{(1)}_{(t+1)j}(I', \widetilde{I}')|\tau]}{\partial \tau} \\
&= \sum_{\boldsymbol{D}} \frac{\partial Pr(\boldsymbol{D}|I_{it} \cup \widetilde{I}_{it}, \tau)}{\partial \tau} \times v^{(1)}_{(t+1)j}(I'(\boldsymbol{D}), \widetilde{I}'(\boldsymbol{D})) \\
&= \sum_{\alpha} Pr(\alpha|I_{it} \cup \widetilde{I}_{it}) \times \sum_{k\in\{11,10\}} \underbrace{\frac{\partial Pr(D_{it}(k) = 1|\alpha, \tau)}{\partial \tau}}_{\text{see Table 1}} \times \underbrace{\left(E[v^{(1)}_{(t+1)j}|\alpha, \text{Paper}] - E[v^{(1)}_{(t+1)j}|\alpha, \text{No Paper}]\right)}_{(*)}
\end{aligned} \tag{7.19}$$

where $(*)$ is the change in the firm's continuation value when employee i produces a paper ($D_{it}(11) + D_{it}(10) = 1$) versus not, expected over other patenting activity, $D_{it}(01)$, which does not vary by τ (Table 1).

The optimal contracts for a new worker maximize (2.14). 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 (2.11,2.15), and the optimal task allocations in (2.12,**??**). In equilibrium, employers set wages and allocate workers to innovation tasks, conditional on information about workers and taking as given the wages set by other employers. The expected labor supply from incumbent and new workers is 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}:

$$V_{1j}\left(\bigcup I_{i1}\right) = \sum_i v_{1j}(I_{i1}) \tag{7.20}$$

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

The FOC for initial wage:

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

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

$$\xi_{i1j} = \frac{b}{\rho_G} \times \left(1 - (1 - \rho_G)s_{j|G} - \rho_G s_j\right) \tag{7.21}$$

The optimal contract can then be written as:

$$w_{i1j} = \left(E_{\alpha|I_{i1}}[Y_j(\alpha, \boldsymbol{\tau})] + \beta\, E[v^{(1)}_{2j}(I', \widetilde{I}')|\tau_{i1j}]\right) \times \xi_{i1j} \times \left(1 + \xi_{i1j}\right)^{-1} \tag{7.22}$$

$$\tau_{i1j} = max\{0, min\{1, \frac{1}{\zeta} E_{\alpha|I_{i1}}[p_\alpha \widetilde{p}_\alpha \phi_j(11) + p_\alpha(1 - \widetilde{p}_\alpha)\phi_j(10) - 1 + \frac{\beta}{\bar{\phi}_j} \times \frac{\partial E[v^{(1)}_{2j}(I)|\tau]}{\partial \tau}]\}\}$$

where wage markdown equals the inverse of labor supply elasticity in (7.21), and the continuation value changes in τ as in (7.19).

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

Disclaimer: I am revising Appendix A2-A4 as of 01/07/2024.

A2. Proof of Propositions 1 and 2

Proof of Proposition 1 - Unique Equilibrium under Monopsonistic Competition

Proposition 1 (Existence and Uniqueness of MPBNE) *In an imperfectly competitive labor market with $b \in (0, \infty)$, $\rho_G \in (0,1)$, and $\lambda_G(\cdot) > 0$, the equilibrium wages $\boldsymbol{w}$ in Definition 1 are unique up to a non-zero scaling factor, and they result in a unique allocation of workers between firms at each possible information state $(I, \widetilde{I})$:*

$$\boldsymbol{s}_{tj}(I,\widetilde{I}) = \begin{cases} \boldsymbol{s}_{1j}\left(\boldsymbol{w}_1(I)\right) & t=1 \\ \boldsymbol{s}^{(1)}_{tj}\left(\boldsymbol{w}_{tj}(I,\widetilde{I}),\, \boldsymbol{w}_{t(-j)}(I)\right) & t>1,\ j=j(i,t-1)\,\text{, as in equation (??)} \\ \boldsymbol{s}^{(0)}_{tj}\left(\boldsymbol{w}_{tj}(I),\, \boldsymbol{w}_{t(-j)}(I,\widetilde{I})\right) & t>1,\ j\neq j(i,t-1)\,\text{, as in equation (??)} \end{cases}$$

In an imperfectly competitive labor market ($\frac{b}{\rho} < \infty$), firms set profit-maximizing wages conditional on the information they have about workers and taking as given the wages set by other firms. Assuming that firms are productive in routine activity $\forall j : f_j > 0$ and there is a positive probability incumbent employees get on the market and look for new jobs $\forall G\, \forall \pi : \lambda_G(\pi) > 0$, employers would set positive wages for all workers, as derived in the backward induction in A1. There exists an equilibrium with wages:

$$\boldsymbol{w}^*_{itj} = \begin{cases} \boldsymbol{w}^{(1)}_{tj}(I_{it} \cup \{\widetilde{y}_{i(t-1)}\}) & j=j(i,t-1)\,\text{, as in equations (2.11, 7.6)} \\ \boldsymbol{w}^{(0)}_{tj}(I_{it}) & j\neq j(i,t-1)\,\text{, as in equations (2.15, 7.11)} \end{cases}$$

In equilibrium, the probability of a worker on the market choosing employer j, as expressed in (7.1), is determined by the wages set by all potential employers:[58]

$$\boldsymbol{p}^*_{ij|C} = \underbrace{\frac{exp\left(\eta_{G(j)}(I_{it}) + \rho_{G(j)} W^*_{iG(j)}\right)}{\sum_{G\in C} exp\left(\eta_G(I_{it}) + \rho_G W^*_{iG}\right)}}_{\text{choose nest } G(j) \text{ in choice set } C} \times \underbrace{\frac{exp\left(b/\rho_{G(j)}\, ln(w^*_{itj})\right)}{exp(W^*_{iG(j)})}}_{\text{choose j within nest } G(j)} \tag{7.23}$$

where the inclusive value for nest G equals $W^*_{iG} := ln\left(\sum_{j\in G} exp(b/\rho_G ln(w^*_{itj})\right)$.

To show that the equilibrium allocation is unique (log wages are unique up to a constant), it would be sufficient to show $M : \mathbf{R}^K \to \mathbf{R}^K$ defined as follows is a

[58]The equilibrium allocation of workers across firms can be viewed as a fixed point of the function $p \circ w$: $p(w(\boldsymbol{p})) = \boldsymbol{p}$.

contraction mapping with modulus less than 1:[59]

$$\forall t\, \forall j > 1:\ M(\boldsymbol{\omega}_{tj}) = \boldsymbol{\omega}_{tj} + p_{tj} - p_{tj}(\boldsymbol{\omega}_t) \tag{7.24}$$

$$\omega_{tj} := \frac{b}{\rho_G(j)} ln(\frac{w_{tj}}{w_{1j}}) \tag{7.25}$$

where ω_{tj} are log wages multiplied by $\frac{b}{\rho} \in (0, \infty)$, relative to that of $j = 1$, and $p_{tj}(\cdot)$ represents the labor supply given wages, which are different for incumbent (**??**) and new workers (**??**).

Following Berry, Levinsohn, and Pakes (1995; henceforth BLP), I show that M satisfies the sufficient conditions for a contraction that are laid out in Theorem 1 of BLP. I focus on the proof for the incumbent workers $\in j$, with labor supply (**??**) at $t > 1$. The proof for new workers is similar. Given any positive wages, the derivatives satisfy:

$$\frac{\partial M_{tj}}{\partial \omega_{tj}} = 1 - \frac{1}{p_{tj}^{(1)}(\omega_t)} \frac{\partial p_{tj}^{(1)}}{\partial \omega_{tj}} = 1 - \underbrace{\frac{\xi_{tj}^{(1)}}{b/\rho_{G(j)}}}_{\text{see (7.5)}} \geq 0 \tag{7.26}$$

$$\frac{\partial M_{tj}}{\partial \omega_{tq}} = -\frac{1}{p_{tj}^{(1)}(\omega_t)} \times \underbrace{\frac{\partial p_{tj}^{(1)}}{\partial \omega_{tq}}}_{\leq 0} \geq 0$$

The cross-derivative depends on if the outside firm $q \in G(j)$:

$$q \in G(j):\ \frac{\partial p_{tj}^{(1)}}{\partial \omega_{tq}} = \lambda_{G(j)} \times \left(p_{j|G(j)} \times p_{q|G(j)}\right) \times E_C[-p_{G(j)|C} + \rho_{G(j)} \times p_{G(j)|C}(1 - p_{G|C})] \tag{7.27}$$

$$q \notin G(j):\ \frac{\partial p_{tj}^{(1)}}{\partial \omega_{tq}} = -\lambda_{G(j)} \times \left(p_{j|G(j)} \times p_{q|G(j)}\right) \times E_C[\rho_{G(j)} \times p_{G(j)|C} \times p_{G(q)|C}]$$

[59] The dimension $K = J \times T \times |\Pi|$, where J: number of firms, T: number of periods, $|\Pi|$: number of beliefs on a grid. Wages w_{tj} are set by employer j at period t for every possible belief $\pi \in \Pi$ on the grid.

The sum of the derivatives in (7.26) for each $j > 1$, at each period $t > 1$:

$$\sum_{q>1} \frac{\partial M_{tj}}{\partial \omega_{tq}} = 1 + \frac{\lambda_{G(j)} \times p_{j|G(j)}}{p_{tj}^{(1)}} \times \Big(E_C[-p_{G(j)|C} \times (1 - \rho_{G(j)} p_{j|C} - (1 - \rho_{G(j)}) p_{j|G(j)})] \tag{7.28}$$

$$+ \sum_{q \in G(j) \setminus \{1,j\}} p_{q|G(j)} \times E_C[p_{G(j)|C} (1 - \rho_{G(j)} + \rho_{G(j)} p_{G(j)|C})]$$

$$+ \sum_{q \notin G(j), q>1} p_{q|G(q)} \times E_C[\rho_{G(j)} \times p_{G(j)|C} \times p_{G(k)|C}] \Big)$$

$$= 1 + \frac{\lambda_{G(j)} \times p_{j|G(j)}}{p_{tj}^{(1)}} \times E_C[\, p_{G(j)|C} \times ((1 - \rho_{G(j)}) \times \underbrace{(1 - 1[G(j) \ni 1] p_{1|C})}_{\leq 1} + \rho_{G(j)} \times \underbrace{\sum_{q>1} p_{q|C}}_{<1} - 1)\,]$$

which satisfies

$$\sum_{q>1} \frac{\partial M_{tj}}{\partial \omega_{tq}} < 1$$

Under the assumption that each firm's routine productivity is positive and bounded, the wages to workers are also positive and bounded. Therefore, M is bounded, satisfying hypotheses (2)(3) in Theorem 1 of BLP.

By Theorem in BLP, we have that M is a contraction mapping of modulus < 1. There is a unique fixed point such that

$$\forall t \, \forall j > 1 : \ \boldsymbol{\omega}_{tj}^* = \boldsymbol{\omega}_{tj}^* + p_{tj} - p_{tj}(\boldsymbol{\omega}_t^*) \tag{7.29}$$

The fixed point $\boldsymbol{\omega}^*$ can be translated to equilibrium wages that are unique up to (nonzero) scaling factor. The equilibrium allocation of workers between firms, as in (7.23) is unique.

□

Special Case: Symmetric Learning under Perfect Competition

Thus far the labor market has been assumed to be imperfectly competitive $\frac{b}{\rho} < \infty$. Suppose that the labor supply is perfectly elastic in each period ($\frac{b}{\rho} \to \infty$ and $\lambda \equiv 1$), and the information is incomplete but symmetric among employers. Once we make such assumptions, the decision to allocate workers to innovation 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 ($\frac{b}{\rho} \to \infty$, $\lambda \equiv 1$) and information is always symmetric, each firm j offers a worker with public information I:*

$$\begin{aligned} \forall t : \boldsymbol{w}_{tj}(I) &= E_{\alpha|I}[Y_j(\alpha, \tau_{tj}(I))] - \zeta(\tau) \\ \textit{in which } \tau_{tj}(I) &= argmax_{\tau\in[0,1]}\, E_{\alpha|I}[Y_j(\alpha, \tau_{tj}(I))] - \zeta(\tau) \end{aligned} \tag{7.30}$$

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; **?**). Otherwise, there is always adverse selection (Greenwald (1986)). It is unclear, however, if incumbent employers would allocate workers to innovation tasks efficiently.

Proof of Proposition 2 - Equilibrium Under Public Information and Perfect Competition

Given $\frac{b}{\rho} \to \infty$ and $\forall G : \lambda_G \equiv 1$, the labor supply elasticity of incumbent and new workers, as expressed in (7.5) and (7.10) both go to infinity. The labor market is perfectly competitive given that the labor supply of every worker is perfectly elastic w.r.t wages.

Plugging $\xi^{(1)}$ into the wage for incumbents at $t = T$, we have $w^{(1)}_{iTj}(\widetilde{\pi}) = MP_j(\widetilde{\pi}, \tau^{(1)}_{iTj})$. Incumbent workers with belief $\widetilde{\pi}$ are paid the full marginal revenue product of labor. Thus, there is no dynamic rent for employers at $(T-1)$. The wage in intermediary periods, as shown in (2.11), also equals a worker's MRPL without leaving any rent to an employer.

Information is assumed to be symmetric between employers. The expectation over $\widetilde{y}$, which indicates the quality of a paper (whether it has a matched patent), can be removed from the wages for new workers as in (2.15). Therefore we have,

$$w_{tj}(\pi) = MP_j(\pi, \tau_{tj}(\pi))$$

for all public belief π that a worker is H-ability.

Since the continuation value equals zero at all employers, allocating workers to innovation tasks also becomes a static decision. The solutions in (**??**,2.12,7.12,7.7,7.22) can be simplified to:

$$\tau_{tj}(\pi) = max\{0, min\{1, \frac{1}{\zeta}(g_j \times q(\pi) - 1)\}\}$$

The costs of innovation tasks are fully deducted from workers' wages (see 2.4). That is, workers are bearing all costs of innovation. They are not credit constrained as they earn a positive wage from routine tasks (under Assumption **??** that $\forall j : M_j > 0$). The choices of innovation 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). □

A3. Model Predictions

Derivation of Prediction 1: Mobility in Response to Public Information
Given information I, denote by $\pi_1 = Pr(H|I \cup \{1\})$ the public belief when a worker has any innovation, and by $\pi_0 = Pr(H|I \cup \{0\})$ the belief otherwise. Assumption 3 implies

$$\pi_1 > \pi_0$$

a) According to (**??**), a worker who produces a public innovation stays at the incumbent employer j with probability

$$p_j^{(1)}(\pi_1) = 1 - \lambda_{G(j)}(\pi_1) \times (1 - E_C[p_{j|C}(\pi_1)]$$

Conditional on common prior, a worker without any new output has labor supply:

$$p_j^{(1)}(\pi_0) = 1 - \lambda_{G(j)}(\pi_0) \times (1 - E_C[p_{j|C}(\pi_0)]$$

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

$$\triangle p_j^{(1)} = p_j^{(1)}(\pi_1) - p_j^{(1)}(\pi_0) = \underbrace{\left(\lambda_{G(j)}(\pi_0) - \lambda_{G(j)}(\pi_1)\right)}_{\leq 0 \text{ under Assumption 2}} \times \left(1 - p_{j|G(j)}(\pi_0) \times E_C[p_{G|C}(\pi_0)]\right)$$
$$+ \lambda_{G(j)}(\pi_1) \times \underbrace{\left(p_{j|G(j)}(\pi_1) \times E_C[p_{G|C}(\pi_1)] - p_{j|G(j)}(\pi_0) \times E_C[p_{G|C}(\pi_0)]\right)}_{\text{choose j again if on market}}$$

Under Assumption 2, $\pi_1 > \pi_0 \rightarrow \lambda(\pi_1) \geq \lambda(\pi_0)$. Unless workers with belief π_1 are much more likely to choose the incumbent j again out of all potential employers, the difference above is negative.[60] It implies workers with a new signal are more likely to leave their incumbent employers than similar coworkers without a signal.

b) Conditional on re-entering the job market, π_1 are more likely to choose firms that are more productive in innovation (higher $g_{j'}$) and can allocate more innovation tasks, relative to the market average. Let j' denote any potential employer, and $\Omega(\pi)$ denote the option value of a worker with belief π on market

$$p_{j'}(\pi_1) = \frac{exp(b \times ln(w_{j'}(\pi_1))}{exp(\Omega(\pi_1))}, \; p_{j'}(\pi_0) = \frac{exp(b \times ln(w_{j'}(\pi_0))}{exp(\Omega(\pi_0))}$$
$$\rightarrow ln\left(\frac{p_{j'}(\pi_1)}{p_{j'}(\pi_0)}\right) = b \times \; ln\left(\frac{w_{j'}(\pi_1)}{w_{j'}(\pi_0)}\right) - \; ln\left(\frac{\Omega(\pi_1)}{\Omega(\pi_0)}\right)$$

Under Assumption 3 and the optimal solutions shown in (2.11, 2.15, 7.6, 7.11), wages are nondecreasing in belief π, resulting in $\Omega(\pi_1) \geq \Omega(\pi_0)$. Moreover, the wage increase is larger at more productive firms (higher $g_{j'}$) that can allocate more innovation tasks to π_1 than other firms on average. In summary, workers with π_1 are more likely to move into j' if the following conditions hold:

(a) $\tau_{j'}(\pi_1) > \tau_{j'}(\pi_0)$;

(b) π_1 is more valuable to j' than to the market average.

The positive assortative matching affects 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 might be able to spend 100% of time on innovation at any firm, and there is no sorting as in a standard AKM framework.[61]

Derivation of Prediction 2: Mobility under Asymmetric Information

[60] The exception with $\left(p_{j|G(j)}(\pi_1) \times E_C[p_{G|C}(\pi_1)] - p_{j|G(j)}(\pi_0) \times E_C[p_{G|C}(\pi_0)]\right) >> 0$ could happen at the most productive firms, where wages increases more in π than at other employers.

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

$$\forall \pi : ln(w_j(\pi)) = \alpha(\pi) + \phi_j \tag{7.31}$$
$$\rightarrow \frac{\Omega(\pi_1)}{\Omega(\pi_0)} = exp(b(\alpha(\pi_1) - \alpha(\pi_0))) \times 1 \tag{7.32}$$
$$ln\left(\frac{p_{j'}(\pi_1)}{p_{j'}(\pi_0)}\right) = b \; (\alpha(\pi_1) - \alpha(\pi_0)) - b \; (\alpha(\pi_1) - \alpha(\pi_0)) = 0 \tag{7.33}$$

Consider two workers $i = 1, 2$ from firm j with a common public belief π at the beginning of period t. The incumbent employer observes $\widetilde{y}_{1(t-1)} = 1 > \widetilde{y}_{2(t-1)}$, while outside employers only observe $y_{1(t-1)} = y_{2(t-1)} = 1$.

a) Denote by $\widetilde{\pi}_{11}$ the private belief about worker 1, and $\widetilde{\pi}_{10}$ the private belief about worker 2. Based on the labor supply in (**??**, **??**), at the beginning of t the difference in the probability a worker stays with the incumbent employer j :

$$\widetilde{\pi}_{11} > \widetilde{\pi}_{10} \longrightarrow p^{(1)}_{tj}(\widetilde{\pi}_{11}, \pi) - p^{(1)}_{tj}(\widetilde{\pi}_{10}, \pi) = \underbrace{\lambda_{G(j)}(\pi)}_{\text{common public belief}} \times \underbrace{\left(p_{tj}(\widetilde{\pi}_{11}, \pi) - p_{tj}(\widetilde{\pi}_{10}, \pi)\right)}_{\geq 0} > 0 \tag{7.34}$$

Given the same public belief π, the two workers are equally likely to get on the market and search for new jobs. The incumbent employer, however, sets a higher wage for the first worker with outputs $(1,1)$ and the second worker with outputs $(1,0)$, as $\widetilde{\pi}_{11} > \widetilde{\pi}_{10}$, resulting in $p_{tj}(\widetilde{\pi}_{11}, \pi) > p_{tj}(\widetilde{\pi}_{10}, \pi)$.

b) Given assumptions on the information structure (**??**, **??**), $\widetilde{y}_{1(t-1)} > \widetilde{y}_{2(t-1)}$ are revealed by $(t+1)$. As the market receives more positive signals about worker 1 than 2, Prediction 1 applies and we have the $(1,1)$ worker more likely to move to a new firm and more productive one than the $(1,0)$ worker.

A4. Model Extension - Forward-looking Workers

So far we have assumed workers consider the utility from wage only, which equals the net present value of a worker-firm match, marked down by the inverse of labor supply elasticity (see 7.5, 7.10). In a more general framework, workers can be forward-looking and take into account their option value in the labor market next period if they enter a firm now. Conditional on wages today, working for a more innovative firm would be more appealing to a high-ability individual who can improve the future market belief about her by producing more innovation today.

For a worker i with market belief π in period t, conditional on her choice of employer $j(i,t)$ there are three potential option values she can reach next period:

1. $\Omega_{i(t+1)}(\pi(1,1))$ if she produces $(y_{it}, \widetilde{y}_{it}) = (1,1)$
2. $\Omega_{i(t+1)}(\pi(1,0))$ if she produces $(y_{it}, \widetilde{y}_{it}) = (1,0)$
3. $\Omega_{i(t+1)}(\pi(1,0))$ if she produces $(y_{it}, \widetilde{y}_{it}) = (0,0)$

We can write her utility of choosing firm j at t given a contract (w, τ) as:

$$\begin{aligned}
u_{itj}(w,\tau;\pi) &= b \times ln(w) + \beta_i \times E_{(y,\widetilde{y})}[\Omega_{i(t+1)}(\pi(y,\widetilde{y}) \mid \tau] + \epsilon_{ijt} \\
&= b \times ln(w) + \beta_i \times \tau \times (\pi \widetilde{h}h + (1-\pi)\widetilde{l}l) \times \Omega_{i(t+1)}(\pi(1,1)) \\
&\quad + \beta_i \times \tau \times (\pi(1-\widetilde{h})h + (1-\pi)(1-\widetilde{l})l) \times \Omega_{i(t+1)}(\pi(1,0)) \\
&\quad + \beta_i \times (1 - \tau \times (\pi h + (1-\pi)l) \times \Omega_{i(t+1)}(\pi(0,0)) + \epsilon_{ijt}
\end{aligned} \tag{7.35}$$

where $\beta_i \in [0,1]$ is the discount factor of workers. Benchmark model assumes $\beta_i = 0$. ϵ_{ijt} are idiosyncratic preferences as before. For simplicity, assume $\epsilon \overset{\text{iid}}{\sim} Gumbel(0,1)$ as in a standard logit model without nested structure.

If belief updating conditional on the innovation outputs are independent of the origin (i.e., $Pr(H|(y,\widetilde{y}), j) \equiv Pr(H|(y,\widetilde{y}))$), then we have:

$$\begin{aligned}
u_{itj}(w,\tau;\pi) &= b \times ln(w) + \gamma(\pi) \times \tau + \epsilon_{ijt} \\
\text{where } \gamma(\pi) =& \beta_i \times (\pi \widetilde{h}h + (1-\pi)\widetilde{l}l) \times \left(\Omega_{i(t+1)}(\pi(1,1)) - \Omega_{i(t+1)}(\pi(0,0))\right) \\
&+ \beta_i(\pi(1-\widetilde{h})h + (1-\pi)(1-\widetilde{l})l) \times \left(\Omega_{i(t+1)}(\pi(1,0)) - \Omega_{i(t+1)}(\pi(0,0))\right)
\end{aligned} \tag{7.36}$$

in which the option value a worker takes into account is reduced to a preference for the allocation to innovation task τ. The preference depends on the current market belief about her only, under the assumption that the belief updating is identical across firms, conditional on τ.

The probability of a new worker choosing firm j in equilibrium, conditional on contract (w, τ) becomes:

$$p_j(w,\tau;\pi) = \frac{exp(b \times ln(w) + \gamma(\pi) \times \tau)}{\sum_{j'} exp(b \times ln(w_{j'}) + \gamma(\pi) \times \tau_{j'})} \tag{7.37}$$

And the optimal task allocation chosen by j solves:

$$\frac{\gamma(\pi)}{u'(w)} + \frac{\partial}{\partial \tau}(MP_j(\pi) - w + \beta E[v_{i(t+1)j}|\tau,\pi]) = 0 \tag{7.38}$$

where the first part is a ratio of the marginal utility of τ vs. wage w, and the second part is the marginal return to spending more time on innovation as in the benchmark model. The benchmark model assumes $\beta_i = 0$, which implies $\gamma(\pi) \equiv 0$, and we are back to the first-order conditions shown in (7.7), for example.

When $\gamma(\pi) > 0$, a worker prefers to spend more time on innovation as it improves her option value in the labor market next period. Equation (7.36) shows that γ is non-decreasing in π, which suggests workers who are more likely to have high-ability further sort themselves into firms allocating more innovation tasks.

To summarize, allowing for forward-looking workers generates additional predictions:

1. Workers with higher market belief π are more likely to choose firms more productive in innovation, all else equal.

2. When $\gamma > 0$, firms can set a lower wage for higher-π workers than in the benchmark where $\gamma = 0$.

These predictions are related to the findings in Stern (2004) that scientists would accept a lower wage to do science. But these results are less relevant for the tradeoff between learning and retention faced by firms I focus on in this paper. τ here represents an amenity that a firm can provide. We will study workers' selection into research jobs in future work.

The main testable predictions on job mobility continue to hold in this model.

B. Data

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).

B1. 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.

B2. Publications Data

The main data source of research papers is Scopus, an abstract and citation databases of peer-reviewed literature launced 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. To maximize matching with an author's employment history, I used the same script that cleans the names of employers on LinkedIn profiles to harmonize the affiliation names from Scopus. We 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;

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. We 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. Some scientometric studies suggest Scopus has better coverage of conference proceedings when compared to Web of Science (e.g., Harzing 2019, Pranckute 2021).

For each paper that is classified as on-the-job research, I recorded the number of citations by year since publication, as well as authors on works that cite this paper to exclude self-citations. Papers with a matched patent application receive more citations over time as shown in Figure 2. The citations on Scopus are mostly conference papers or journal articles. In future work, I will look at citations between papers and patents.

B3. Match between Papers and Patent Applications

I collected patent data from the 2022 release of the Patent Examination Research Dataset (PatEx), which contains publicly viewable patent applications from the Public Patent Application Information Retrieval System (Public PAIR) as of June 2023.[62] For each patent application, I collected the names of inventors, and related parent/child application within a family, dates of the earliest filing, publication of the application, and grant date if a patent is eventually granted. We then merged patent applications with USPTO's Patent Assignment Dataset to obtain the assignee of an application, which are typically the employer(s) of inventors.[63]

Before matching with research papers, I cleaned the names of authors and

[62]PatEx 2022 "contains detailed information on more than 13 million publicly-viewable provisional and non-provisional patent applications to the USPTO and over 1 million Patent Cooperation Treaty (PCT) applications. It is based on data that OCE downloaded from the Patent Examination Data System (PEDS) in June, 2023."https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-examination-research-dataset-public-pair

[63]Patent Assignment Dataset 2021 contains "detailed information on 9.6 million patent assignments and other transactions recorded at the USPTO since 1970 and involving roughly 16.5 million patents and patent applications. It is derived from the recording of patent transfers by parties with the USPTO." https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-assignment-dataset

assignees, using the same scripts for cleaning the names of authors and affiliations from Scopus. To reduce computational burden, I focus on papers with at least one Ph.D. author for whom I have collected a LinkedIn profile. The matching is done in two steps:

1. For each (paper, author) pair in year t, I looked for all (patent app, inventor) with the inventor = author that are initially filed between years $[t-3, t+3]$. Considering the number of authors/inventors matched at the paper/patent level, I drop matches if:
    - Less than half of the inventors on a patent application are matched, and less than half of the authors on a paper are matched.
    - The number of inventors on a potential matched patent is $< 1/3$ or > 3 the number of authors on the paper.
2. Merge the matched (paper, patent, author/inventor) from (1) with author affiliations from Scopus at (paper, author) level, and with assignees at (patent, assignee) level.
    - Keep (paper, author/inventor, patent) matches if the author's affiliation is matched with one of the patent assignees.

The matching by authors and affiliations above generate about 439,000 potential matches at (paper, patent, author) level, which span between about 75,000 papers and 84,000 patent applications.

To further enhance match quality, I compare the titles and abstracts of papers from Scopus, with titles and abstracts for potentially matched patent applications, which are extracted from Google Patents Public Datasets via BigQuery. We used OpenAI's Ada V2 text embedding model to create numerical representations of paper or patent abstracts.[64] Each embedding is a vector of dimension 1,536. The more similar a patent abstract to a paper's, the smaller the distance between their vector embeddings. This measure of paper-patent similarity is available for 85% of the potential matches.

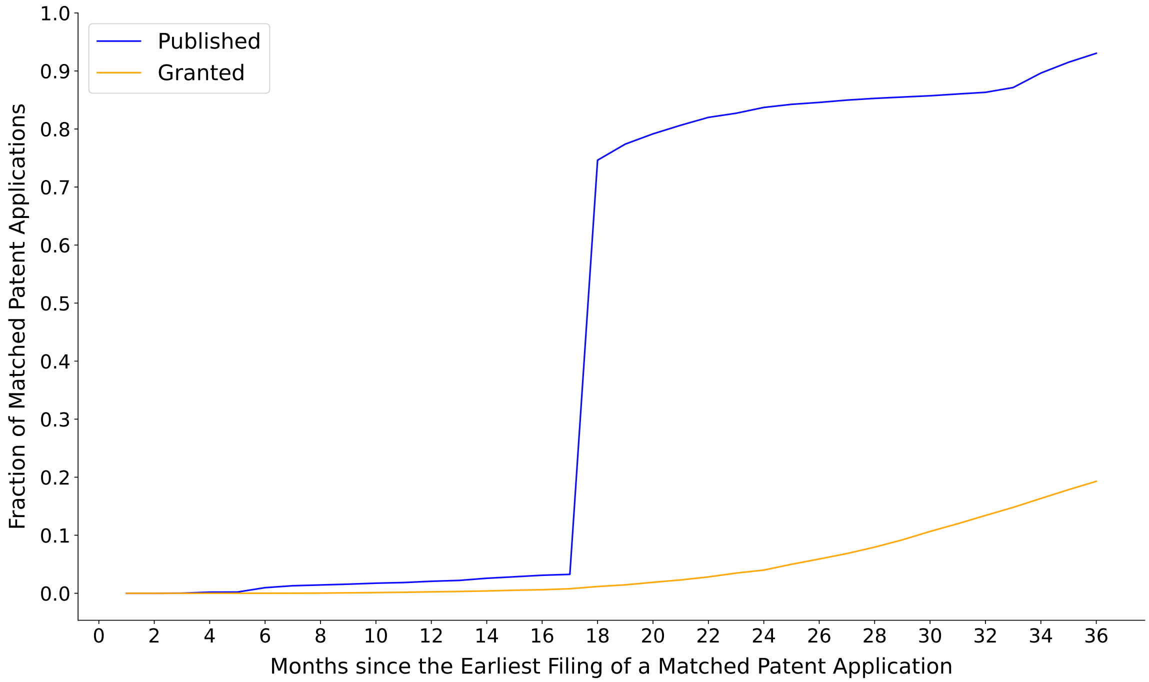
For each CS paper, I sort the potentially matched patent applications as follows and select the first one as the best possible match:

1. # matched authors, # matched inventors on a patent in descending order;
2. at least one author affiliation can be matched with patent assignee;
3. prefers patent application filed in t, the year a paper is published;
4. distance between text embeddings, in ascending order;

[64] Ada V2 outperforms Google's BERT and OpenAI's earlier embedding models (Neelakantan et al. 2022).

5. prefers patent applications filed in t, then $t-1$, then $t+1$.

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).

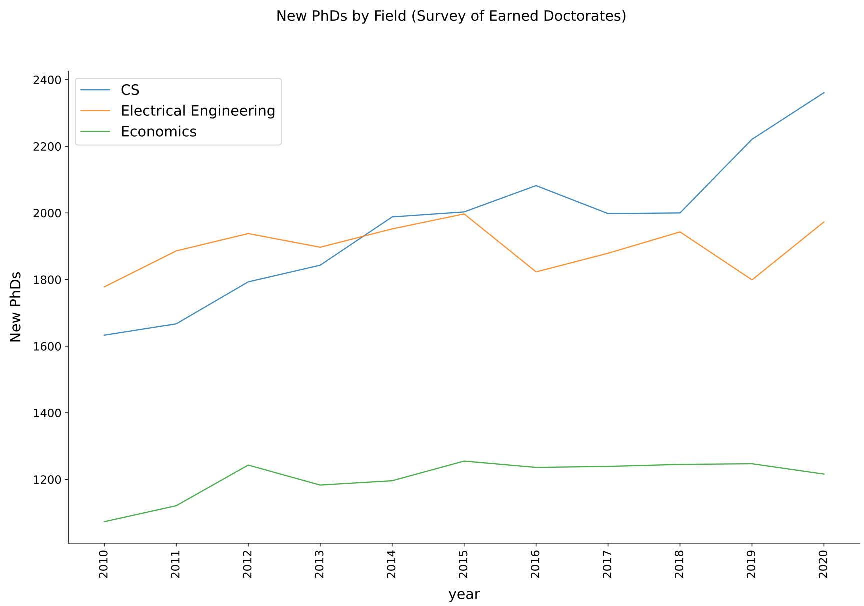
Preferred qualifications:

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

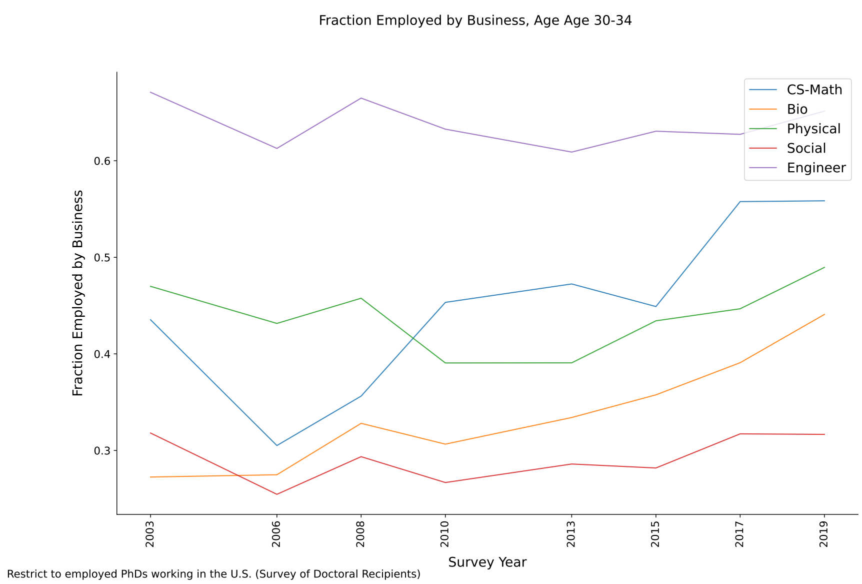
Notes: This figure shows recent postings of research scientist jobs at Amazon and Google. Both ads explicitly indicate a graduate degree in computer science as a basic qualification for this type of jobs, and list "publication records" as preferred qualifications.

Figure B3: CS PhDs in NSF Surveys

(a) New PhDs (Survey of Earned Doctorates)

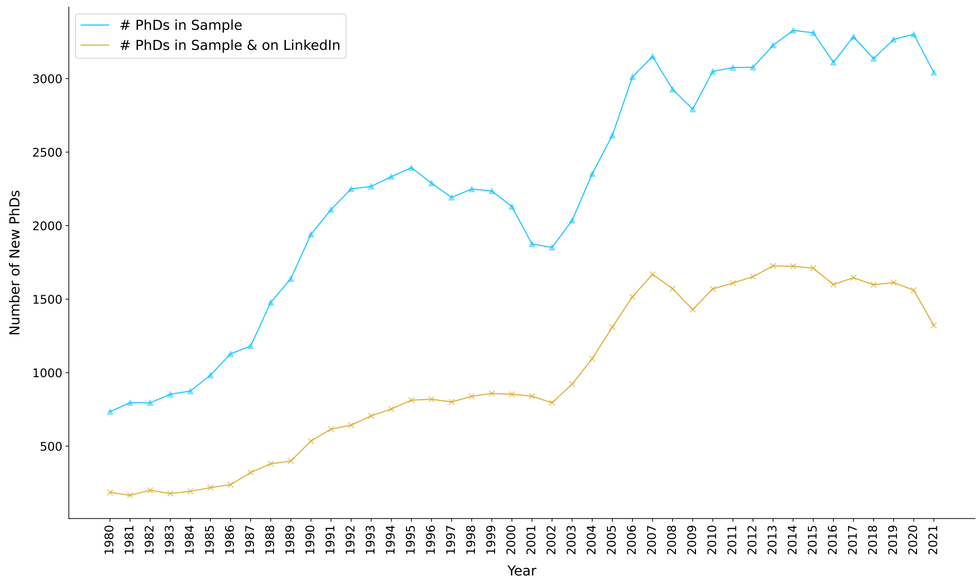


(b) Fraction Employed by Business, Age 30-34



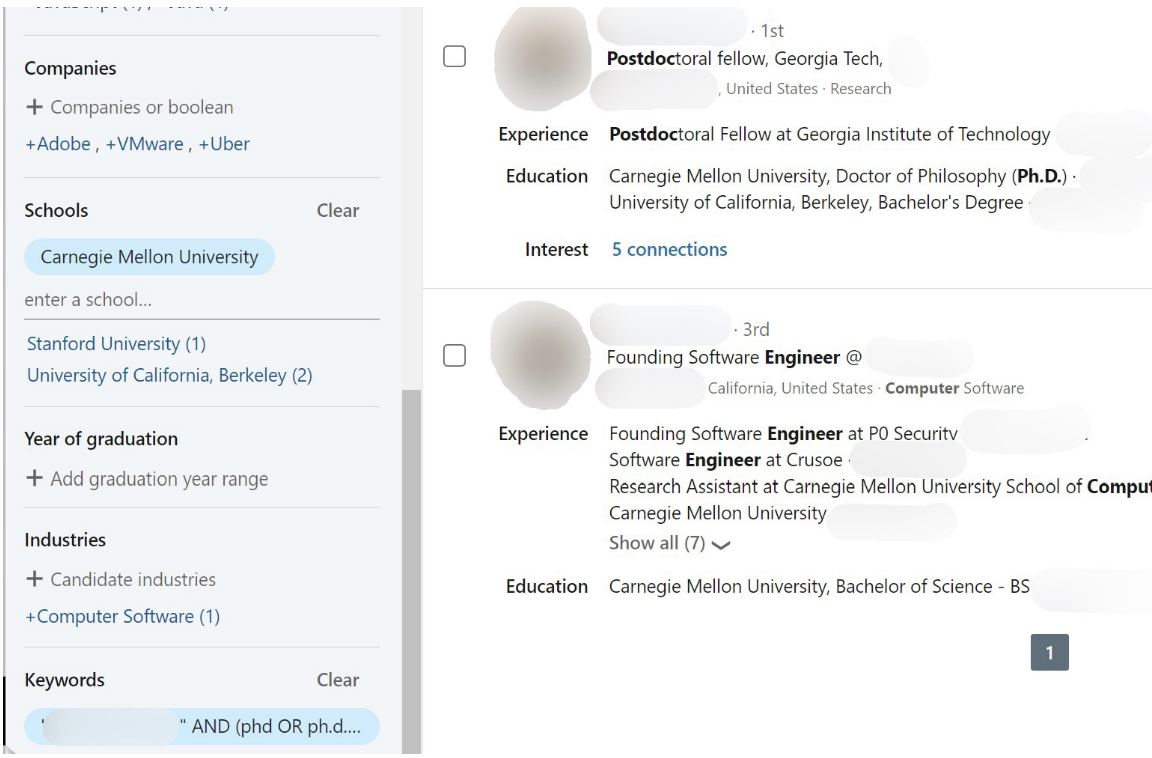
Notes: (a) displays the number of new PhDs in the Survey of Earned Doctorates by NSF. (b) come from the the Survey of Doctoral Recipients, restricted to Ph.D. recipients in the U.S. with nonmissing employer information between age 30-34.

Figure B4: Number of PhD Dissertations and Matched LinkedIn Profiles by Graduation Year



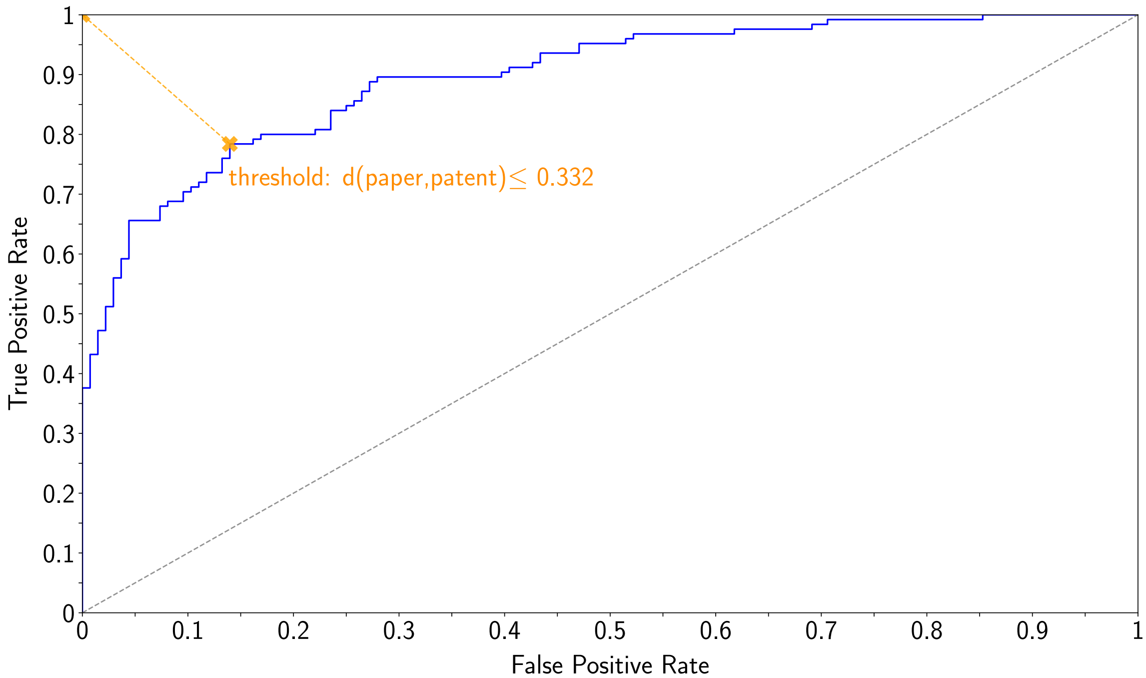
Notes: The blue line (top) shows the number of Ph.D. recipients in Computer Science or Electrical Engineering identified in ProQuest dissertation database or various school-specific sources (Appendix Table B2) by graduation year from 1980 to 2021. The yellow line plots the number of Ph.D.s who are matched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

Figure B5: LinkedIn Platform



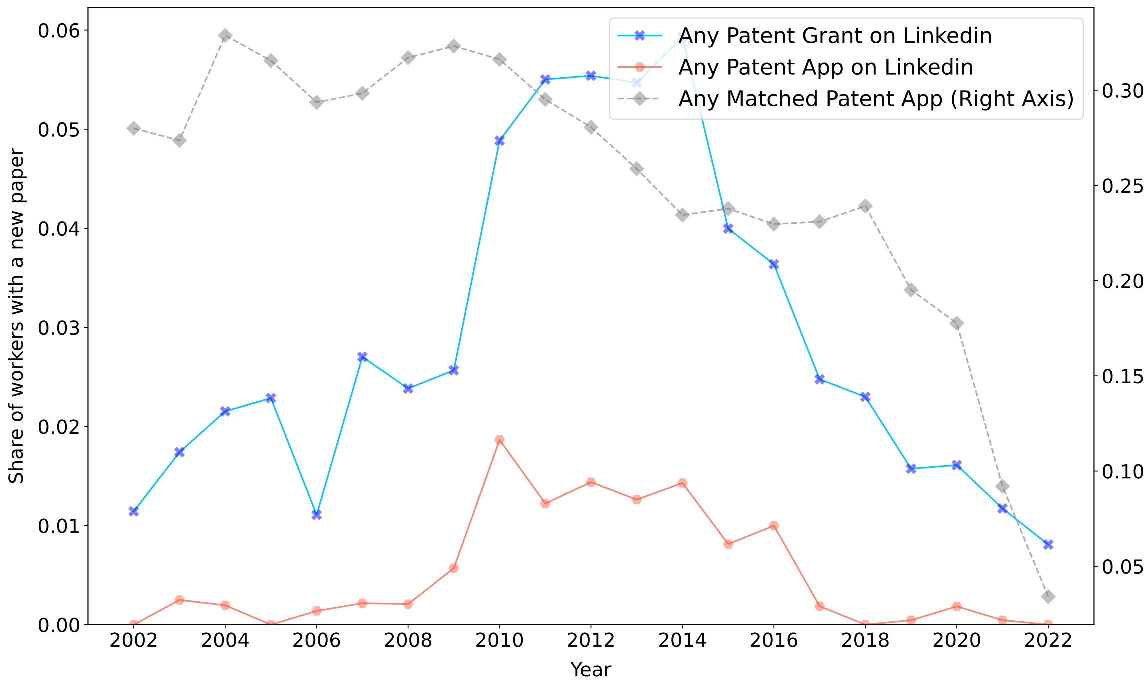
Notes: This figure shows the outputs of one query on the LinkedIn Recruiter Lite platform. The query includes the full name of a CS Ph.D. and keywords about a "Ph.D." degree and about CS such as "computer science" or "electrical engineering". The search is also restricted to CMU, where the person receives 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 $(0, 1)$.

Figure B7: Patents on LinkedIn



Note 1: This figure shows the time series of the share of workers who listed a patent on LinkedIn by year, conditional on having a new paper that year. The patents section on a LinkedIn profile may include either a patent grant or application, and provides the grant and/or filing date(s). The blue line (left axis) shows the share of workers who have a new paper in a given year (based on publication records) and list a granted patent the same year on LinkedIn. The red line (left axis) shows the share of workers who have a new paper and list a patent application the same year. The gray line (right axis) shows the share of workers who also have a patent application matched to a new paper, for comparison.

Note 2: Patents (applications) listed on LinkedIn may not correspond to the ones that can be matched to a paper. This plot, however, suggests workers are much more likely to advertise their granted patents rather than applications, especially in more recent years when the applications are yet to be published by USPTO.

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

year	i_pro	i_add	ld	ld_out	ld_matched
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)

	ProQuest Dissertations			School-specific Sources		
School	# 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.

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

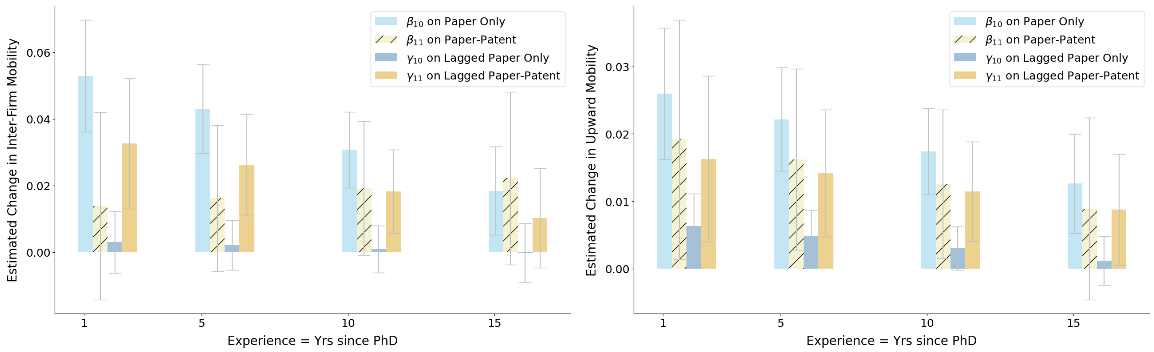
Firm	Team Overlap	Text Distance	Papers: Title	Papers: M/Yr	Matched Patent Applications: Title	Filing M/Yr	Published M/Yr
Microsoft	100%	0.247	FROID OPTIMIZATION OF IMPERATIVE PROGRAMS IN A RELATIONAL DATABASE	12/2017	METHOD FOR OPTIMIZATION OF IMPERATIVE CODE EXECUTING INSIDE A RELATIONAL DATABASE ENGINE	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: We add interactions between $D_{it}(10), D_{it}(11)$, Lagged-$D_{it}(10)$, Lagged-$D_{it}(11)$ and years of experience since PhD to 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 $k = 11, 10$, 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 : $\mathbf{D_{it}(10)}$ vs. $D_{it}(11)$						
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 $[\mathbf{t-3, t-1}]$: Lagged-$D_{it}(10)$ vs. Lagged-$D_{it}(11)$						
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						
$D_{it}(01)$	-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-$D_{it}(01)$	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 $k = 11, 10$ 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 : $\mathbf{D_{it}(10)}$ vs. $D_{it}(11)$						
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 $[\mathbf{t-3, t-1}]$: Lagged-$D_{it}(10)$ vs. Lagged-$D_{it}(11)$						
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						
$D_{it}(01)$	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-$D_{it}(01)$	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 : $\mathbf{D_{it}(10)}$ vs. $\boldsymbol{D_{it}}$(11)							
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 $\mathbf{[t-3, t-1]}$: Lagged-$\boldsymbol{D_{it}}$(10) vs. Lagged-$\boldsymbol{D_{it}}$(11)							
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 $k = 11, 10$, 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 : $\mathbf{D_{it}}$(10) vs. D_{it}(11)									
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 $[\mathbf{t-3, t-1}]$: Lagged-D_{it}(10) vs. Lagged-D_{it}(11)									
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. Estimation

D1. Details on Estimation

Disclaimer: I am revising this appendix as of 12/27/2024.

D2. Additional Estimation Results

Appendix Figure D: Change in Publication Rate in the Absence of Employer Learning

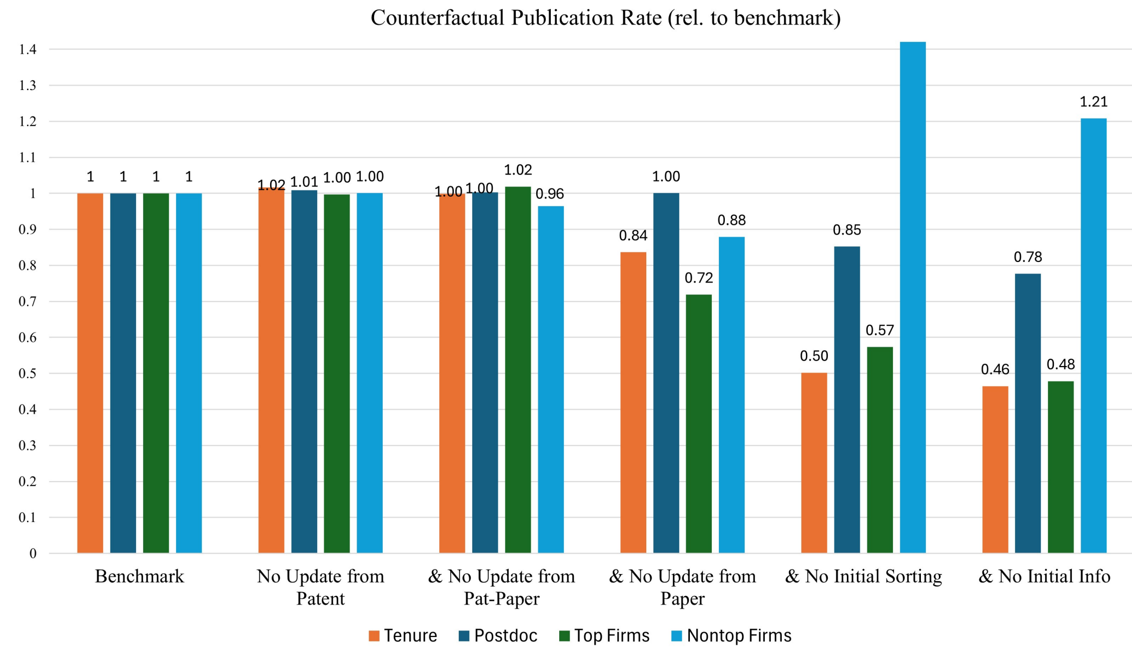


Table D1: Model Parameters

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 (2.7)		0.63
ρ_G	$1-$ corr. of ϵ_{itj} for $j \in$ nest G	$\rho_1 = 1$ for postdoc	$(0.78, 0.45, 0.88)$ at $G \neq 1$
$(\eta_{1,G}, \eta_{2,G})$	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 (2.6) $\lambda_G(\pi) = \lambda_{0,G} \times (1 + \lambda_{1,G} \times \pi)$, at $t > 1$	$(0.40, -0.50)$ at $G = 1$	$(0.04, -0.5)$ at $G = 0$ $(0.08, 0.10)$ at $G = 2$, $(0.13, 0.99)$ at $G = 3$
$(\Lambda_{AJ}, \Lambda_{JA})$	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
$\phi_j(10)$	j's proportional return to paper		Table D2
$\phi_j(01)$	j's proportional return to patent	j-fixed effect in patenting	Table D2
$\phi_j(11)$	j's proportional return to paper-patent	$\phi_j(11) = 1.25 \times \phi_j(10) + 0.25 \times \phi_j(01)$	
ζ	cost of innovation: $c(\pi, \tau) = \frac{\zeta}{2}\tau^2$	0.30	
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: The 16 employers (Table D2) belongs to four nests: Tenure Track ($G = 0$), Postdoc ($G = 1$), Top Firms in Industry ($G = 2$), and Nontop Firms in Industry ($G = 3$). There are 56 parameters that are estimated by maximizing the joint likelihood of job movements and innovation outputs (5.2), using the limited-memory BFGS optimization algorithm (**?**). See Section 5.1 for estimation details. Additional assumptions are fully specified in Appendix D.

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

			Returns to Innovation	
j	Description	Baseline $\overline{\phi}_j$	Paper $\phi_j(10)$	Patent $\phi_j(01)$
Nest 0. Academia - Tenure Track				
0	Nontop Schools	0.298	0.485	0.080
1	Top 25 CS	0.011	0.800	0.092
Nest 1. Academia - Postdoc				
2	Postdoc at Nontop Schools	0.015	0.412	0.097
3	Postdoc at Top 25 CS	0.008	0.476	0.096
Nest 2. Industry - Top Firms				
4	IBM	0.005	0.490	0.533
5	Microsoft	0.022	0.365	0.257
6	Amazon	0.019	0.182	0.223
7	Facebook (Meta)	0.021	0.255	0.193
8	Apple	0.018	0.087	0.283
9	Google (Alphabet)	0.060	0.253	0.197
Nest 3. Industry - Nontop Firms (Grouped by Patenting FE)				
10	Above 90th Percentile	0.087	0.293	0.425
11	80th-90th	0.214	0.273	0.220
12	70th-80th	0.088	0.236	0.121
13	50th-70th	0.171	0.224	0.082
14	25th-50th	0.157	0.263	0.046
15	<25th Percentile	0.122	0.228	0.001

Notes: I classify the 7,000 unique employers into 16 groups (indexed by j), which belong to four nests (G). This table displays the maximum-likelihood estimates of the baseline productivity, $\overline{\phi}_j$, and their returns to CS papers, $\phi_j(10)$. The productivity in patenting, $\phi_j(01)$, is calibrated based on the estimated j fixed effect in a regression of patent application on firm fixed effects, conditional on worker characteristics. I further calibrate the return to a paper with a matched patent as $\phi_j(11) = 1.25 \times \phi_j(10) + 0.25 \times \phi_j(01)$.. In academia ($G \in \{0, 1\}$), "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 sorted by the regression estimate for j fixed effect in patenting, conditional on worker characteristics and time trend.

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.