ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 214/26

Employer Learning, Sorting, and Productivity: Evidence from Computer Scientists

Alice Wu

www.rfberlin.com AUGUST 2026

Employer Learning, Sorting, and Productivity: Evidence from Computer Scientists

Authors

Alice Wu

Reference

JEL Codes: J24, J42, J62

Keywords: Learning, Signal, Search, Sorting, Productivity

Recommended Citation: Alice Wu (2026): Employer Learning, Sorting, and Productivity: Evidence from Computer Scientists. RFBerlin Discussion Paper No. 214/26

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

Employer Learning, Sorting, and Productivity: Evidence from Computer Scientists

Alice H. Wu*

August 6, 2026

Abstract

How does employer learning affect the allocation of talent and aggregate productivity? I study this question in the labor market for computer science (CS) Ph.D.s, using the job histories and post-Ph.D. publications of 31,000 graduates from 2000 to 2021. Publishing a CS conference paper raises the probability that a researcher moves to a top tech firm in the following year, especially for workers who start at less productive firms in industry. The increase in upward mobility upon publication is larger for less experienced workers and for scarcer signals, such as first authorship or papers with a matched patent. These patterns are consistent with predictions from an equilibrium search model with public employer learning. Estimating the model, I find that learning from post-Ph.D. publications accounts for 14% of overall CS publications, as it reallocates high-ability researchers to employers that provide more opportunities to publish. This contribution exceeds that of initial sorting, underscoring that much of worker ability is revealed on the job rather than at entry.

*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 Chris Taber, John Kennan, Rasmus Lentz, Chao Fu, Mandy Pallais, Jesse Shapiro, David Cutler, Louis Kaplow, Francesca Miserocchi, Antonio Coran, Omeed Maghzian, Jacob Weber, Ayushi Narayan, Audrey Tiew, Ariel Gomez, four anonymous scientists from industry, and seminar participants for helpful discussions and comments. 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 how labor is allocated across firms. 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; Lange 2007). Existing tests of employer learning, however, rely on indirect correlates of ability (Pinkston 2009; Schönberg 2007). The public signals that learning models assume employers observe are rarely visible to researchers. This missing-data challenge makes it difficult to measure employer learning directly, let alone to quantify its consequences for the efficiency of talent allocation that theoretical frameworks emphasize (e.g., Terviö 2009; Waldman 1984).

This paper asks how much employer learning matters for sorting and productivity. I address the missing-data challenge by building a new dataset that combines the employment histories of Ph.D.s in computer science (CS) with their publications in major conference proceedings. Comparing similar coworkers, I show how publications produced on the job affect inter-firm mobility. I build an equilibrium search model with public employer learning, which incorporates information frictions into sequential auction frameworks (Postel–Vinay and Robin 2002; Bagger, Fontaine, Postel-Vinay, and Robin 2014). The model guides my empirical tests for employer learning and allows me to quantify structurally the impact of learning on aggregate publication productivity.

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 keep publishing at academic conferences.[2] These conferences are peer-reviewed and, in most CS fields, matter more than journals (Ernst 2013; Vrettas and Sanderson 2015). Roughly half of papers at major CS conferences since 2000 involve at least one Ph.D. author from the industry. Initial information from the Ph.D. degree is less predictive of future research success in CS than in economics.[3] Publications after the Ph.D. reveal additional information about research ability, and to my knowledge, this is the first paper to use them to test

[1] The number is based on the Survey of Earned Doctorates by the National Science Foundation (Online Appendix Figure OA.1). 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.

[2] According to the Survey of Doctoral Recipients (Figure OA.1), the share of new CS Ph.D.s entering the industry has been increasing over the past 20 years and has exceeded 50% since 2017.

[3] Using the data on economists in Sarsons (2017), I find that Ph.D. school and cohort fixed effects can explain about 18.8% of the variation in publications before tenure in economics. The same characteristics can explain only 5.5% of the variation in publications five years after a CS Ph.D.

directly for employer learning.

My empirical analysis builds on a new equilibrium on-the-job search model with information frictions about worker ability. Firms differ in productivity and workers differ in research ability, which is unknown to all. As in sequential auctions (Postel–Vinay and Robin 2002), workers and firms are forward-looking, and a worker accepts an outside offer when the expected match value with the new employer exceeds that with the current employer. The key departure from existing search models is employer learning. Given the same opportunities to work on research, higher-ability workers publish more, and firms observe the publications and update the market belief accordingly. Belief updates, however, do not map one-to-one into wages: two workers with the same belief can be paid differently due to their bargaining positions. Learning can instead be measured through job-to-job transitions. If offer arrivals increase with market belief, the equilibrium features positive assortative matching between workers and firms. The structural estimates support this assumption: offers from top industry firms increase with market belief, as do offers from academia to postdocs and workers in industry. Positive assortative matching implies that workers who publish are more likely to move up the job ladder — the basis for my test of public learning.

I combine Ph.D. dissertations, public LinkedIn profiles, and post-Ph.D. research output for 31,000 computer scientists who graduated between 2000 and 2021. I test for public employer learning by comparing the job mobility of workers who produce a paper with similar coworkers who do not. Any publication this year predicts higher job-to-job (J2J) mobility across firms within one year, conditional on experience, position characteristics, education, and firm-year fixed effects — comparing coworkers at the same firm and time. The effect is strongest for workers in industry but outside the top firms. I define top firms throughout as {Alphabet (Google), Meta (Facebook), Amazon, Apple, Microsoft, IBM} — the "big tech" firms over the 2000–2022 sample period. These firms are also among the most valuable employers based on revealed preference from J2J flows (Page, Brin, Motwani, and Winograd 1999; Sorkin 2018) — the first five are exactly the five highest-valued, and IBM ranks thirteenth.[4]

For workers at non-top firms in the industry, publishing a paper today predicts a 3.9 percentage points (p.p.) higher probability of moving to a different employer and a 2.1 p.p.

[4]Despite ranking below other top firms according to the PageRank algorithm, IBM has a strong research division that employs about 3,000 researchers and was still the second-most-valuable tech company in 2011 (Macrotrends LLC 2026; Raghavan, Khare, and Gambetta 2025).

higher probability of moving to a top firm within one year — nearly doubling the mean upward mobility rate of coworkers. An event study around the first publication after the Ph.D. shows a 3.3 p.p. rise in moves to top firms within a year, and the increase continues through year five. These patterns provide strong evidence for public employer learning from publications as positive signals about worker ability.

For workers already at top firms, publishing does not raise retention, but it predicts moving to a research-scientist role within top firms rather than an engineering one. Academics are also more likely to move to top firms upon publishing, suggesting that top firms recruit high-ability researchers from both industry and academia (Miserocchi, Noray, and Wu 2026).

The benchmark model also predicts that mobility responses increase with the scarcity of the signal. I test this with two types of scarce signals: (1) whether the worker is listed as first author, and (2) whether the publication has a matched patent application. First authorship raises upward mobility from non-top to top firms by a further 1 p.p., relative to coworkers who publish but are not first authors. The impact of first authorship on sorting to top firms is more immediate and persistent for workers from academia, where author order is more informative of intellectual contribution rather than management hierarchy.

The second scarce signal is a matched patent application. About 25% of papers by industry researchers are accompanied by a patent application filed by their employers. Such papers are more highly cited in later years, suggesting higher quality. Workers whose first publication after the Ph.D. has a matched patent show a steeper rise from non-top to top firms than coworkers with a paper alone, though the gap between the two groups is not statistically significant. The gap is significant when research output is measured over the past three years instead: authors who publish a paper with a matched patent in that window are 1.0 p.p. more likely to move to a top firm the next year, whereas those with papers alone show a smaller 0.3 p.p. increase.

The lagged mobility response to the signal of research ability provided by a combined paper and patent application is partly driven by the delay in information about patenting. By default, the United States Patent and Trademark Office (USPTO) publishes patent applications 18 months after their initial filing.[5] The presence of a matched patent application is therefore revealed later than the paper itself, offering a test for asymmetric learning between the incumbent and outside employers.[6] Authors of papers with a matched patent are

[5] Title 35 U.S.C. 122 (AIPA 1999).

[6] Workers rarely advertise undisclosed patent applications on LinkedIn, seemingly respecting non-

less likely to move than other authors within a year, when only the paper is public. But in three years, once most patent applications are disclosed, they are more likely to leave their original employers and enter top firms. An event study around the USPTO disclosure date itself, however, shows no jump in mobility. The evidence is therefore mixed: outside employers do not appear to react to the patent disclosure per se; recognizing high-quality work may simply take them longer than it takes the incumbent.

Given the robust evidence for employer learning from publications, I focus on public learning in the structural analysis and ask how much it matters for aggregate publication productivity, as a measure of productive efficiency. Answering this requires linking the sorting of heterogeneous workers across firms to publication output. Motivated by the reduced-form analysis, I extend the benchmark model to consider three employer groups — academia, non-top firms, and top firms — each representing a separate job ladder and providing workers with different opportunities to publish. Key model parameters include the within-group and cross-group offer arrival rates, which vary with employer belief. Publication productivity depends not only on worker ability but also on employer group, so the sorting of workers across groups matters for overall publication output.

The model is estimated by the simulated method of moments, matching publication output, upward and downward mobility within each group, and transitions across the three groups. The estimated model reproduces the empirical relationship between publications and job-to-job transitions. The main counterfactual shuts off belief updating from publications; I compare its impact on productivity with that of initial sorting and initial observables, computing each channel's average marginal contribution following Shapley (1953). Employer learning from post-Ph.D. publications has the highest Shapley value: it raises mean publications per worker-year by 14%, led by a 20% increase in academia and a 17% increase at top firms. These gains come from a larger share of high-ability workers reaching top firms and academia later in their careers.

The second-most-important mechanism is initial sorting, which allows the first placement to reflect information observed by employers at hiring but not by econometricians. It accounts for 12% of overall publication output, but matters less for publication and job transitions as workers gain experience. This dynamic path is consistent with Altonji and Pierret (2001), who show that information available at labor market entry becomes less correlated with earnings as employers learn.

This paper makes two main contributions. First, I contribute to the employer learning

disclosure agreements.

literature by providing evidence of public employer learning from CS researchers' career trajectories after publications. 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 by firms) over time to employer learning. The underlying model of these studies posits that employers update their beliefs when new signals arrive, but these signals are rarely observable except from within-firm personnel records (Kahn and Lange 2014). In the CS labor market, I observe publications that are produced on the job and signal workers' research ability. By relaxing the perfect-competition assumption often made in employer learning models, I derive predictions on changes in job mobility following publications from the model equilibrium, and use them to test for public learning even in the absence of individual wage data.

Second, this paper provides a tractable model that considers search frictions and information frictions simultaneously. Most search models assume that worker ability is known, with Bagger et al. (2014) as an exception in that they have idiosyncratic productivity shocks. Pinkston (2009) considers asymmetric learning in sequential auctions and derives pure bidding strategies as in Milgrom and Weber (1982) when outside firms hold private information of their own. Without such information, outside bidders may adopt mixed strategies as in Hendricks and Porter (1988) and Li (2013). However, these papers are restricted to common-value auctions, and once firms differ in productivity, it is unclear how wages are set. Relaxing the common-value assumption is important in this setting, where high-ability researchers sort to top firms and academia. I therefore take one step forward by introducing symmetric employer learning into a sequential auction model. Under information symmetry and efficient contracting, the steady-state equilibrium remains tractable and accounts for the evolution of employer beliefs. I use the estimated model to quantify the role of employer learning in sorting and productivity, an exercise that can be extended to other labor markets.

2 An Equilibrium Search Model with Employer Learning

I lay out a model of the sorting of workers to firms in a labor market with both search frictions and learning about worker ability. Workers are engaged in on-the-job search as in Postel–Vinay and Robin (2002), Bagger and Lentz (2019), and Bagger, Fontaine, Postel-Vinay, and Robin (2014). Employers update their beliefs over time, conditional on public signals, as in Kahn (2013); Kahn and Lange (2014); Schönberg (2007). In the steady-

state equilibrium under public learning, I characterize the sorting between heterogeneous workers and firms and derive testable predictions on worker mobility upon the revelation of productivity signals.

2.1 Environment and Timing

The labor market is populated by rational, risk-neutral, and forward-looking workers and firms, who share a discount rate $\rho > 0$. Time is discrete with an infinite horizon.

Workers. There is a unit mass of workers in every period. When workers enter the labor market, they are endowed with a time-invariant research ability $\theta \in [0,1]$ that is unknown to all agents, including the worker herself. The labor market only knows that θ is drawn from a Beta distribution with shape parameters (α_0, β_0); the common prior is therefore Beta(α_0, β_0).[7]

Firms. Firms vary by productivity $\phi \in [\underline{\phi}, \overline{\phi}]$ and have an additively separable production technology across workers. Firms provide each employee N independent opportunities to publish each period. Letting Y denote the number of publications a worker produces in a period, the distribution of Y conditional on research ability θ is Binomial(N, θ). The flow value of a match between a firm with productivity ϕ and a worker with ability θ is:

$$m(\phi, \theta) = \phi \cdot (1 + E[Y|N, \theta] \cdot \eta) \tag{1}$$

where $\eta > 0$ is the additional return per publication, proportional to ϕ. In the benchmark model, N is common across employers; Section 5.1 presents an extended model with employer heterogeneity in N. I estimate that academic employers and top firms in industry provide more opportunities to publish than other employers, so the allocation of workers across employers matters for overall publication productivity.

Information Set. I focus on symmetric learning, in which all agents observe the same information. Workers are characterized by the cumulative number of publications since labor market entry, denoted by S, and the number of years they have been employed, t. Under the assumptions of a common prior and common N across employers, the pair (S,t) is a sufficient statistic for the market belief about worker ability. Labor market entrants start off with $S = 0$, $t = 0$, and the common prior Beta(α_0, β_0). Under Bayesian updating from the binomial-distributed signals, the posterior belief for workers in state (S,t) is Beta$(\alpha_0 +$

[7] The mean of the prior belief is $\frac{\alpha_0}{\alpha_0+\beta_0}$ and the variance is $\frac{\alpha_0\beta_0}{(\alpha_0+\beta_0)^2(\alpha_0+\beta_0+1)}$. Holding the mean fixed, the variance is decreasing in $\alpha_0 + \beta_0$: the sum of the shape parameters measures the precision of the belief.

$S, \beta_0 + t \cdot N - S$).[8] Following Bagger et al. (2014), I assume that the state remains unchanged for the unemployed, whereas for the employed, real experience increases from t to $t+1$ and S accumulates with realized publications.

Offer Arrivals. Job offers are drawn from a continuous distribution F with support $[\underline{\phi}, \overline{\phi}]$, density f, and complementary CDF $\overline{F}(\phi)$. An unemployed worker draws an offer with probability $\lambda_0 \in (0,1)$. An employed worker receives an outside offer with probability $\lambda_1(\pi) \in (0,1)$, where π is the posterior mean of the employer belief and equals $E[\theta|S,t] = \frac{\alpha_0+S}{\alpha_0+\beta_0+N \cdot t}$ for those in state (S,t). The dependence on π can be derived from a model with endogenous search intensity (Bagger and Lentz, 2019): workers with a more favorable belief have stronger incentives to search to match with more productive firms.[9] Bagger et al. (2014) similarly allow λ_1 to increase with worker ability to generate positive assortative matching. I characterize the equilibrium sorting under the assumption that λ_1 is increasing in the posterior mean, which is supported by the structural estimates (Section 5).

Timing of Events. At the beginning of each period, a measure μ of entrants with experience $t = 0$ is randomly matched with firms drawn from F; μ equals the probability that an experienced worker permanently exits the labor market, so the labor force is stationary. For workers with at least one year of experience, each period unfolds as follows:

1. *Production and signals.* The employed produce at their employers, and on-the-job publications $y \in \{0, 1, \ldots, N\}$ are realized and publicly observed; the state moves from (S,t) to $(S+y, t+1)$, where S denotes cumulative publications before this period. The unemployed do not publish and the state remains unchanged.

2. *Shocks.* The employed face mutually exclusive shocks: permanent exit (μ), layoff into unemployment (δ), displacement with a new employer drawn directly from F (κ), and the arrival of an outside offer ($\lambda_1(E[\theta|S+y, t+1])$).[10] The unemployed exit with probability μ or draw an offer with probability λ_0, with $\mu + \lambda_0 < 1$.

3. Job transitions are realized, and the next period begins.

[8] The Beta-Bernoulli or Beta-Binomial update scheme has a wide range of applications. Katz (1985) uses it to model the possibility of recall after layoffs. More recently, Lim (2025) uses it to represent college students' learning about their STEM abilities from course grades.

[9] Online Appendix B.3 presents a richer model with endogenous search intensity.

[10] The shocks are mutually exclusive draws from a single lottery, with $\mu + \delta + \kappa + \sup \lambda_1 < 1$. The κ shock introduces involuntary job-to-job transitions down the job ladder; Bagger and Lentz (2019) refer to it as an advance layoff notice.

2.2 Value Functions

In the steady state, wage contracts are renegotiated sequentially upon mutual agreement (Postel–Vinay and Robin 2002). Following Bagger and Lentz (2019) and Lindenlaub and Postel-Vinay (2023), I focus on the joint value of a match $V(\phi,S,t)$, where ϕ is firm productivity and (S,t) summarizes the state of the worker. Under efficient contracting, a new match is formed only if it yields a joint value exceeding that of the current match.

Wages are set by generalized Nash bargaining, and workers have bargaining power $\gamma \in (0,1)$. Outside offers from inferior firms may increase the wage but do not change the joint value.[11] The worker leaves firm ϕ upon receiving an offer from firm ϕ' that yields a higher joint value, i.e., $V(\phi',S,t) > V(\phi,S,t)$. She moves to ϕ' and uses the current match value $V(\phi,S,t)$ as her bargaining position.

The bargaining process also applies to the unemployed, whose outside option is the value of unemployment $U(S,t)$ when negotiating with any arriving firm. The value of an unemployed worker in state (S,t) is:

$$U(S,t) = b + \frac{\lambda_0 \gamma}{1+\rho} \int_{\underline{\phi}}^{\overline{\phi}} \big(V(\phi',S,t) - U(S,t)\big)\, dF(\phi') + \frac{1-\mu}{1+\rho} U(S,t) \tag{2}$$

in which b denotes unemployment benefits, and the value of exiting the labor market is normalized to zero. Unemployment benefits are assumed to be low enough that $V(\underline{\phi},S,t) \geq U(S,t)$ in all states (S,t), so the unemployed accept any arriving offer.

The joint value between a worker in state (S,t) and a firm of productivity ϕ in steady state is the sum of the expected flow output (1) and the next-period continuation value, discounted by factor $1+\rho$. Relative to existing on-the-job search models, the key innovation is to allow market beliefs about worker ability to evolve based on public signals. The continuation value, accounting for all scenarios described in the timing of events, is a weighted average over realized signals $y \in \{0,1,\ldots,N\}$, with marginal likelihood $Pr(Y=y|S,t)$ as weights, given employer belief $\theta \sim \text{Beta}(\alpha_0+S, \beta_0+t\cdot N-S)$.[12] Outside offers arrive

[11] Meeting a less productive firm may improve the worker's bargaining position and generate a higher wage (Cahuc, Postel-Vinay, and Robin 2006; Bagger et al. 2014). Wage setting does not affect mobility under efficient contracting: the worker at ϕ can capture $V(\phi',S,t) + \gamma \cdot (V(\phi,S,t) - V(\phi',S,t))$ when an offer from ϕ' with $V(\phi',S,t) < V(\phi,S,t)$ arrives. She does not leave the original firm, and the joint value remains $V(\phi,S,t)$. Lentz, Maibom, and Moen (2026), for example, simply assume that meetings with inferior firms do not lead to a wage increase.

[12] Given belief $\theta \sim \text{Beta}(\alpha_0+S, \beta_0+t\cdot N-S)$ and $Y|\theta \sim \text{Binomial}(N,\theta)$, the probability that the worker produces y publications in a period is $Pr(Y=y|S,t) = \binom{N}{y} \frac{B(\alpha_0+S+y, \beta_0+(t+1)\cdot N-S-y)}{B(\alpha_0+S, \beta_0+t\cdot N-S)}$, where the beta function $B(a,b) \coloneqq \int_0^1 \theta^{a-1}(1-\theta)^{b-1} d\theta$.

with probability $\lambda_1(E[\theta|S+y,t+1])$, evaluated at the posterior mean.

$$
\begin{aligned}
& V(\phi,S,t) \qquad (3)\\
&= \underbrace{m(\phi,E[\theta|S,t])}_{\text{flow}} + \sum_{y=0}^{N} Pr(Y=y|S,t)\,\frac{\delta+\kappa}{1+\rho}\,\underbrace{U(S+y,t+1)}_{\text{unemployed}}\\
&\quad + \sum_{y=0}^{N} Pr(Y=y|S,t)\,\frac{\kappa}{1+\rho}\cdot \underbrace{\gamma\int_{\underline{\phi}}^{\overline{\phi}} \big(V(\phi',S+y,t+1)-U(S+y,t+1)\big)\,dF(\phi')}_{\text{surplus from involuntary J2J}}\\
&\quad + \sum_{y=0}^{N} Pr(Y=y|S,t)\,\frac{\lambda_1(E[\theta|S+y,t+1])}{1+\rho}\cdot \underbrace{\gamma\int_{\phi}^{\overline{\phi}} \big(V(\phi',S+y,t+1)-V(\phi,S+y,t+1)\big)\,dF(\phi')}_{\text{surplus from bargaining with a higher-value firm}}\\
&\quad + \sum_{y=0}^{N} Pr(Y=y|S,t)\,\frac{1-\mu-\delta-\kappa}{1+\rho}\,\underbrace{V(\phi,S+y,t+1)}_{\text{match value at }\phi}.
\end{aligned}
$$

The joint value is strictly increasing in firm productivity ϕ (Lemma 1 in Appendix B.1). Hence workers accept an outside offer if and only if it comes from a more productive firm.

2.3 Steady-State Equilibrium

The steady-state distributions of workers across firms are derived from flow-balance equations. Workers are characterized by ability θ, cumulative publications S, and real experience t. Denote by $u(\theta,S,t)$ the density of unemployed workers with ability θ in public state (S,t), and by $l(\theta,S,t)$ the density of the employed.

Entrants are characterized by $(S,t)=(0,0)$, with ability drawn from the population distribution Beta(α_0,β_0); because they are assumed to start with a job, $u(\theta,0,0)=0$ and $l(\theta,0,0)\propto\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}$. At experience level $t>0$, the flow-balance equation for the employed with public state (S,t) and true ability θ is:

$$
l(\theta,S,t)=\lambda_0\,u(\theta,S,t)+(1-\mu-\delta)\sum_{y=0}^{\min(N,S)} Pr(Y=y|\theta)\,l(\theta,S-y,t-1) \qquad (4)
$$

On the left-hand side, all the currently employed in state (S,t) constitute the outflow, as experience increments to $t+1$. The inflow on the right-hand side comprises the unemployed whose state remains unchanged at (S,t), and the employed previously in state $(S-y,t-1)$, which is updated to (S,t) upon producing y publications. By solving the system of flow-balance conditions recursively for the unemployed and the employed, I show that the ratio

$\frac{l(\theta,S,t)}{l(\theta,0,0)}$ equals the probability of accumulating S publications in tN trials times the probability of remaining in the labor force through t years of experience (Lemma 2 in Online Appendix B.1).

The joint distribution of workers and firms in steady state depends on inflows from prior beliefs, introducing a recursive structure that links the joint distribution across public states. Denote by $G(\phi|\theta,S,t)$ the distribution of firms conditional on worker ability θ and public state (S,t). I show that conditional on the public state, $G(\phi|\theta,S,t)$ is independent of ability θ (Lemma 3) and the following holds:

$$G(\phi|S,t) \propto F(\phi)\cdot\left(\frac{\lambda_0\delta}{\mu+\lambda_0}+\kappa\right) + (1-\widetilde{\delta}-\lambda_1(E[\theta|S,t])\cdot\overline{F}(\phi))\sum_{y=0}^{\min(N,S)}\frac{\binom{N}{y}\binom{(t-1)N}{S-y}}{\binom{tN}{S}}G(\phi|S-y,t-1). \tag{5}$$

The first term on the right-hand side, proportional to sampling distribution $F(\phi)$, captures inflows from unemployment and involuntary J2J moves. The second term captures the flows of the employed from prior $(S-y,t-1)$ to posterior (S,t), accounting for exogenous separation risks ($\widetilde{\delta}:=\mu+\delta+\kappa$) and poaching by more productive firms, which occurs with probability $\lambda_1(E[\theta|S,t])\cdot\overline{F}(\phi)$.

At any experience $t\geq 1$, $G(\phi|S,t)<F(\phi)$ for all $\phi\in(\underline{\phi},\overline{\phi})$: workers are more concentrated at productive firms than the offer distribution would suggest, a result that is well established in equilibrium search models (e.g., Postel–Vinay and Robin 2002). If λ_1 is constant, the conditional distribution is independent of the cumulative signals S, and hence of the employer belief. Proposition 1 establishes positive assortative matching under the assumption that λ_1 is increasing in the posterior mean.

Proposition 1 (Positive Assortative Matching) *Assume the offer arrival rate λ_1 is increasing in the posterior mean. For all $\phi\in(\underline{\phi},\overline{\phi})$, at any experience $t\geq 1$, the following holds:*

a) Given cumulative signals $S,S'\in\{0,1,\ldots,tN\}$,

$$S>S'\Rightarrow G(\phi|S,t)<G(\phi|S',t). \tag{6}$$

b) Let $G(\phi|\theta,t)$ denote the distribution of firms conditional on true ability θ and experience t. For any $\theta,\theta'\in(0,1)$,

$$\theta>\theta'\Rightarrow G(\phi|\theta,t)<G(\phi|\theta',t). \tag{7}$$

Proof in Appendix B.2. This equilibrium result establishes positive assortative matching between firms and workers (a) with more cumulative signals S, and (b) with higher underlying ability θ, both in the first-order stochastic dominance sense. This analytical result under employer learning complements Bagger et al. (2014), who show PAM numerically when offer arrival rates increase with human capital, and Lindenlaub and Postel-Vinay (2023), who establish PAM via multi-dimensional matching.

The assumption that λ_1 increases with the posterior mean $E[\theta|S,t]$ can be rationalized by endogenous search intensity as in Bagger and Lentz (2019) and Lentz (2010): given the supermodular production function (1), workers with a higher employer belief gain more from moving to more productive firms and therefore search harder. Alternatively, firm productivity could be two-dimensional: workers who publish more are more likely to cross belief thresholds and enter firms that are more productive in research, as in assignment models (Gibbons and Waldman 1999; Gibbons and Katz 1992). I sketch both ideas in Appendix B.3. The structural estimates in Section 5 show that job arrival rates from top firms strongly increase with the posterior employer belief, further supporting this assumption.

2.4 Predictions on Job Mobility

The steady-state sorting results in Proposition 1 have direct implications for worker mobility across firms following signal realization, which I formalize as testable predictions below. The predictions are derived under the assumption that λ_1 is differentiable and increasing in the posterior mean ($\lambda_1' > 0$). The last two predictions additionally require λ_1 to be locally linear.[13] The proofs are in Appendix B.4.

Prediction 1 (Job Mobility in Response to Publications) *Relative to coworkers with the same prior, workers who publish more exhibit higher mobility across firms and are more likely to move to more productive firms.*

Prediction 1 follows directly from the steady-state worker flows and the assumption that λ_1 increases with the posterior mean. Consider coworkers who share the same employer belief Beta(α,β) and the same opportunities to publish N. The posterior mean given y new

[13] Based on the structural estimates in Section 5, the 95-th percentile of posterior mean is 0.40, and the semi-elasticity of offer arrivals from top firms is 1.36 (equation 12). For $\pi \in [0, 0.40]$, the exponential factor $\exp(1.36\,\pi)$ is nearly linear in π: a linear approximation has slope 1.80 and $R^2 = 0.995$, with a RMSE less than 0.015 against a total variation of 0.72. Local linearity is therefore a good approximation at the estimated arrival rates.

publications is $\frac{\alpha+y}{\alpha+\beta+N}$, which is increasing in y. Under the assumption $\lambda_1' > 0$, workers who produce more publications face a higher arrival rate of outside offers and are more likely to move to more productive firms, as long as the current firm satisfies $\phi < \overline{\phi}$.

Prediction 2 (Belief Precision) *Given a current belief Beta(α,β), the update in the posterior mean based on publications is smaller for workers with a more precise belief, measured by $\alpha+\beta$. If λ_1 is locally linear, the smaller belief update translates to a smaller mobility response to publications.*

Consider two workers who share the same current belief and the same N, and who produce y and y' publications, respectively. The gap in their posterior means, $\frac{|y-y'|}{\alpha+\beta+N}$, is smaller when the belief precision $\alpha+\beta$ is higher.[14] If λ_1 is increasing and locally linear, a smaller change in the posterior mean translates into a smaller change in the arrival rate of outside offers, and hence a smaller mobility response. Since the belief precision $\alpha+\beta$ rises with experience as opportunities to publish accrue, this prediction implies that the mobility response to a publication is stronger for less experienced workers. It can also be tested by comparing workers whose ability is assessed more precisely at entry, such as graduates of well-known schools versus others.

Prediction 3 (Signal Scarcity) *The update in the posterior mean based on publications is larger when there are fewer opportunities to produce such signals, i.e., when N is lower. If λ_1 is locally linear, the larger belief update translates into a larger mobility response to signals.*

When there are fewer opportunities to publish, each new publication generates a larger change in the posterior mean. In the benchmark model, N is common across employers. I relax this in the structural model and estimate that there are more opportunities to publish in academia than in industry, and at top firms than at non-top ones. This prediction can be tested by comparing the mobility responses to publications across these origins, and by exploring rarer signals such as first-authored papers or publications with a matched patent.

3 Data

This section describes the data I collected on the career trajectories and research outputs of Ph.D. computer scientists. I discuss the matching between Ph.D. dissertations and public

[14] Holding the posterior mean fixed, a larger $\alpha+\beta$ implies a smaller belief variance $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

LinkedIn profiles, and measures of on-the-job research based on conference publications and patent applications.

3.1 Ph.D. Graduates in CS Fields

I focus on Ph.D. graduates in CS and 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.[15] The share of new CS Ph.D.s entering industry as opposed to academia has been increasing over the past two decades and has been over 50% since 2017. On the ProQuest Theses and Dissertation Database, I identified about 68,000 Ph.D. dissertations in computer science or electrical engineering from the top 60 CS schools in the United States, between 2000 and 2021.[16] Each dissertation record includes the full name of the doctoral recipient, institution, and year of Ph.D.

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

I developed a program that queries public profiles on LinkedIn Recruiter Lite, a subscription-based recruiting platform that provides access to LinkedIn's member database with advanced search filers. For each person in the dissertation sample, I submitted a query by the person's full name, Ph.D. institution, and degree information.[17] About 51% of the queries returned at least one LinkedIn profile, resulting in 31,000 fully matched profiles for graduates between 2000-2021.[18]

Each profile is formatted as a résumé, containing public information on education and employment history. I constructed a longitudinal dataset of post-Ph.D. job history for the fully matched individuals. On average, a CS Ph.D. has 1.9 industry employers, 0.2 postdoc employers, and 0.3 tenure-track employers after Ph.D. (Online Appendix Table OA.1). For each worker$\times$year observation, I identify the primary employer and the corresponding job

[15] See Online Appendix Figure OA.2 for research scientist job ads. CS Ph.D.s may also work as engineers, but they often start as senior software engineers or research engineers, the latter of whom can also publish.

[16] I use the ranking of CS departments on CSRankings. Appendix Tables OC.1-OC.2 display the number of dissertations by year and by institution.

[17] With a Recruiter Lite subscription, I searched for a small batch of graduates each time and viewed the returned profiles on the platform. Appendix Figure OC.1 shows a sample query.

[18] A profile is considered matched to the graduate only if the full name and Ph.D. institution match exactly, and the year of Ph.D. is the same if available on the profile. The matching rate is higher for cohorts after 2006 (Figure OC.2). Differences between matched and unmatched graduates are small: the unmatched publish 0.8 more papers on average (3.6% relative to the mean) and have 0.8 p.p. higher share of academic affiliations (a 2% difference).

title.[19] A job-to-job transition is defined as a change in one's primary employer from one year to the next. About 13% of workers at non-top firms transition to a new employer per year, whereas workers at top firms or in academia are less mobile (Table OA.2).

3.3 On-the-job Research

To measure the research output of Ph.D. computer scientists, I collected papers published in 80 CS conferences and two machine learning journals, all of which are used by CSRankings to rank CS departments. Unlike in economics, conference proceedings are the primary publication venue in computer science for most subfields and are widely recognized as a key criterion for academic promotion (Computing Research Association 1999; Amato, Daescu, Gini, Hayes, Larson, Li, Lin, Lopresti, Siek, Spafford, and Zorn 2025). I searched for papers at each conference/journal×year on Scopus, a large-scale publication database produced by Elsevier.[20] Each publication record includes a complete list of authors and their affiliations, which indicate the employer of an author at the time of publication.

To match papers with education and job histories, I cleaned and harmonized institution names from all three sources: author affiliations on CS publications, Ph.D. institutions on ProQuest, and employers on LinkedIn. A paper is labeled as pre-Ph.D. research if it was published before graduation and the author affiliation matches her graduate school. After Ph.D., a paper is considered on-the-job research if the author affiliation matches her employer at the time of publication.[21] About 30% of matched computer scientists have at least one on-the-job research publication after the Ph.D. (Table OA.1). At the person-year level, the publication rate is higher at top firms: 11% of employees at top firms publish at least one paper per year, versus 3% at non-top firms (Table OA.2).

4 Empirical Evidence of Learning and Sorting

Using publication and job records of CS Ph.D. graduates, I show that on-the-job publications predict higher upward mobility across a range of outcomes, providing evidence

[19] If there are multiple employers in a year, I rank them in descending order by (1) full-time position (over contract or visiting), (2) months worked at the employer during the year, and (3) tenure at the employer.

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

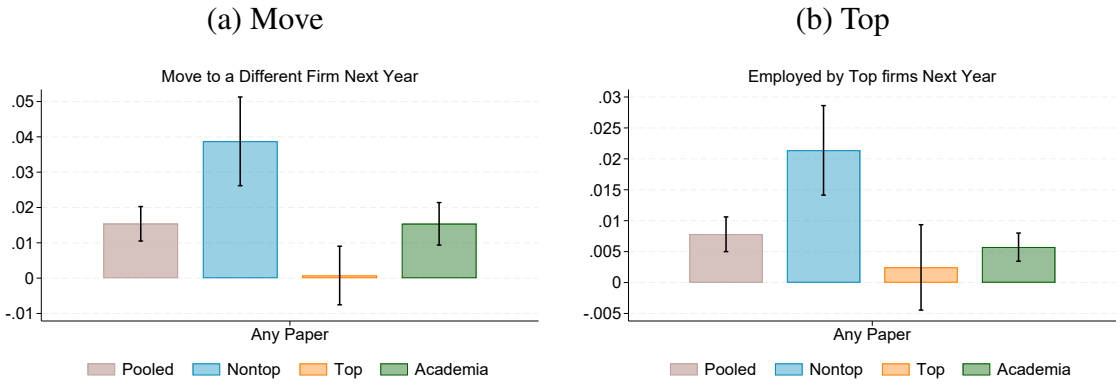
[21] The publication cycle is substantially shorter in computer science, making it unlikely for a dissertation chapter to be published as conference proceedings years later. I further check whether coauthors on a paper are affiliated with the Ph.D. school or with the current employer. Roughly 1% of post-Ph.D.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

for the model predictions on public employer learning. I then exploit patents matched to publications to test for asymmetric learning; the evidence is weaker than that for public learning.

4.1 Job Mobility in response to Publications

Prediction 1 states that workers who publish are more likely to move, and to move up, relative to similar coworkers. I test this prediction by regressing year-to-year job transitions on an indicator for any publication this year, conditional on experience after the Ph.D., current position characteristics, demographics, Ph.D. school×cohort, and firm×year fixed effects.

FIGURE 1. Any publication on job outcomes next year



Notes: This figure shows the predictive effect of any publication this year on (a) moving to a different employer, and (b) moving to a top firm in the next year. Each regression controls for a quartic in experience since the Ph.D. (divided by 10), indicators for current position characteristics (academic, engineer, scientist, manager, and senior), gender, and region of bachelor's degree, with Ph.D. school×cohort fixed effects and firm×year fixed effects. Each panel reports the pooled estimate in the full sample and then the estimates separately by whether workers are at non-top firms, at top firms, or in academia this year. 95% confidence intervals are computed from robust standard errors clustered at the Ph.D. school×cohort level.

Workers who publish this year are 1.5 p.p. more likely to change employers the next year than similar coworkers, according to the pooled estimate in panel (a) of Figure 1. I then estimate the regression separately by worker origin. Those who publish a paper while working for a non-top firm in industry are 3.9 p.p. more likely to move. The effect is also positive for academics. But there is no effect for workers already at top firms, consistent with the model, in which workers at the top have no more productive firms to move to.

Panel (b) shows that workers who publish are not moving randomly. Any publication predicts a 2.1 p.p. increase in moving to a top firm the next year for workers at non-top firms, and a 0.6 p.p. increase for academics, suggesting that top firms also draw productive researchers from academia, not just from other firms in industry.

These mobility responses also hold at the intensive margin. Appendix Figure OA.3 shows similar findings when publication output is measured by the asinh of publications at the worker-year level. The contrast between workers publishing at non-top firms and top firms can also be driven by a larger belief update at places with fewer opportunities to publish, in support of Prediction 3: non-top firms are estimated to provide one to three fewer opportunities to publish than top firms per year in Section 5.

4.1.1 Upward Mobility from Non-top to Top Firms

Workers at non-top firms show the strongest association between publications and job mobility. I therefore estimate how upward mobility to top firms evolves around the first publication one to three years after the Ph.D., using a dynamic difference-in-differences model:[22]

$$M_{it} = \sum_{l \neq 0} \beta_l D_{it}^{(l)} + \sum_{l \neq 0} \gamma_l D_{it}^{(l)} \cdot \text{Published}_i + \alpha_i + \mu_t + \varepsilon_{it} \tag{8}$$

where $\text{Published}_i = 1$ for workers who publish a paper at event time $l = 0$. The control group comprises workers without a publication in the first three years after the Ph.D., for each of whom I draw a placebo event time from an ordered logit, following Kleven, Landais, and Søgaard (2019).[23] The event dummy $D_{it}^{(l)} = 1$ when calendar time t is l periods from the first (real or placebo) publication. The model is estimated on workers employed by non-top firms at $l = 0$, with worker fixed effects, which also absorb origin-firm effects. The identifying assumption is that, in the absence of the publication signal, treated workers would have followed career trajectories parallel to those of control workers, whose placebo event times are drawn conditional on initial characteristics.

Panel (a) of Figure 2 shows the dynamic effects of the first publication on employment by top firms at the annual level. There is no differential pre-trend between workers with and without a publication within three years of the Ph.D. One year after the first publication,

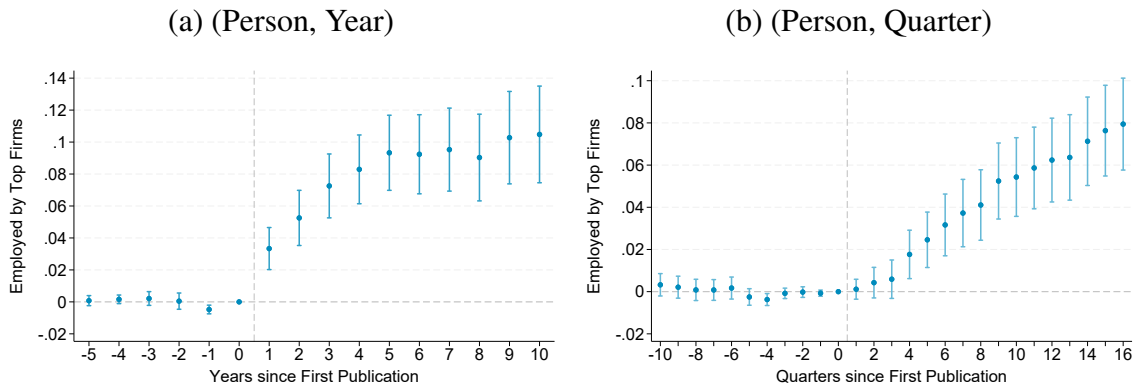
[22] I drop publications in the Ph.D. year, which are typically dissertations rather than on-the-job research.

[23] Among workers with a publication within three years, I estimate an ordered logit regression of years to the first publication on characteristics of the first job, and Ph.D. school×cohort fixed effects. For workers without a publication in three years, I draw a placebo event time (1, 2, or 3 years after the Ph.D.) from the fitted categorical probabilities of this model.

authors are 3.3 p.p. more likely to be employed by a top firm. Upward mobility continues to rise over the following years and stabilizes around 9 p.p. by year five. Zooming in on quarterly data, panel (b) shows that transitions to top firms begin four quarters after the publication and continue to rise for about four years.[24] This profile is consistent with the search model: beliefs update when the paper is published, but outside offers arrive at finite rates, so transitions to top firms accumulate over time rather than jump on impact. These results hold when I replace worker fixed effects with origin-firm fixed effects, comparing publishing and non-publishing coworkers (Figure OA.4).

The event study thus provides further evidence that workers who publish at non-top firms experience a substantial and persistent increase in upward mobility.

FIGURE 2. Employment by top firms after the first publication



Notes: This figure shows the difference-in-differences estimates of $\{\gamma_l\}$ in equation (8) — the dynamic effects of the first publication on employment by top firms. All workers are employed by non-top firms at event time $l = 0$ and have a job record in every period between -1 and 3. The treated publish their first paper 1–3 years after the Ph.D. while employed by non-top firms at 0. The controls publish none within three years of the Ph.D. and are assigned a placebo event time from an ordered logit on characteristics of the first job and Ph.D. school×cohort fixed effects. Panel (a) is annual, and panel (b) quarterly. Standard errors are robust and clustered at the worker level.

4.1.2 Retention at Top Firms and Other Mobility Outcomes

Publication is not associated with overall retention at top firms. Decomposed by role, however, the picture changes: the first two panels of Appendix Figure OA.5 show that employees of top firms who publish are 3.5 p.p. more likely to become research scientists,

[24] I use the conference date to identify the quarter of publication. For the control group, I assign a random quarter in the placebo event year.

and 0.7 p.p. less likely to become engineers, than comparable coworkers. Publications thus predict which role a worker holds within a top firm, not whether she stays. Vesting stock options may also encourage workers to switch roles internally rather than leave the firm.

The scientist-versus-engineer contrast also holds for workers outside top firms, who after publishing are more likely to enter top firms as research scientists than as engineers — further evidence that publications help top firms identify and channel high-ability researchers into research roles.

Using wage records from H-1B visa and green card petitions for foreign workers, I define a move to a higher-wage firm as an alternative measure of upward mobility. Panel (c) of Figure OA.5 mirrors the pattern in Figure 1: workers who publish at non-top firms are 2.3 p.p. more likely to enter a higher-wage employer the next year. Excluding moves to top firms (which often pay higher wages), panel (d) shows that the effect falls below 0.8 p.p. for workers at non-top firms and below 0.5 p.p. for academics. This contrast confirms that upward mobility to higher-wage firms operates mostly through transitions to top firms.

The last two outcomes in Figure OA.5 concern promotions, defined as a move to a more senior job title or to a higher-paid position based on posted wages for foreign workers. Any publication predicts a 2.5 p.p. increase in promotion to a senior job title for workers at non-top firms, and a 1.8 p.p. increase for workers at top firms (panel (e)). For moves to a higher-wage position, panel (f) shows that the effect of any publication exceeds 3 p.p. for workers at non-top firms and academics, but is near zero at top firms. Because posted wages miss pay variation within firm$\times$position, this near-zero effect at top firms need not indicate an absence of advancement: combined with the results on becoming a scientist versus an engineer (panels (a)–(b)), it is plausible that workers who publish at top firms advance their careers by transferring to a research role.

Together, these results support Prediction 1: workers who publish outside the most productive firms are more likely to move, and to move up to top firms. Even within top firms, publication predicts internal shifts towards research roles.

4.2 Heterogeneity Analysis

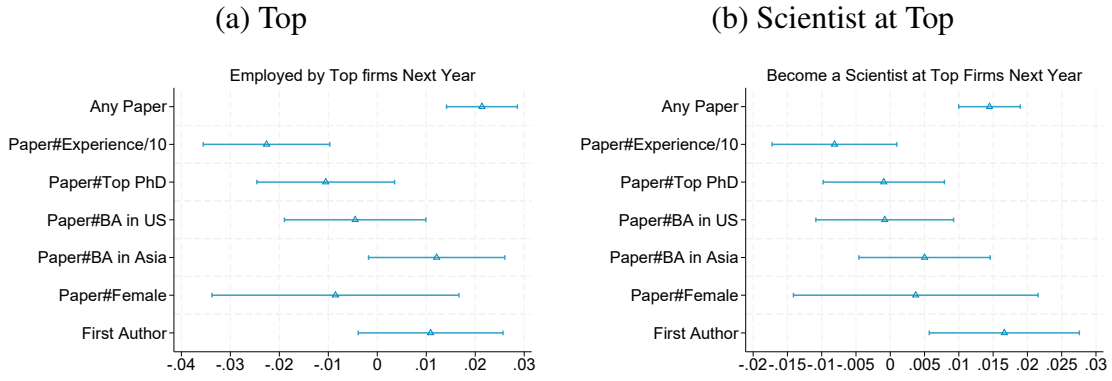
4.2.1 Heterogeneity by Worker Characteristics

Prediction 2 states that the mobility response to publication output is smaller for workers with more precise beliefs. To test this prediction, I consider several characteristics that are correlated with the precision of employer beliefs. For each characteristic C, I regress next

year's employment outcome on any publication this year, C, and their interaction, using the same controls and firm×year fixed effects as described under Figure 1.

Figure 3 plots the interaction coefficients for two measures of upward mobility from non-top firms: panel (a) for employment by a top firm next year, and panel (b) for employment as a research scientist at a top firm.

FIGURE 3. Heterogeneity in Mobility Response by Worker Characteristics



Notes: The sample comprises workers currently at non-top firms. The first row reports the baseline effect of any publication this year on (a) employment at top firms and (b) employment as a research scientist at a top firm next year, using the specification described under Figure 1. Each subsequent row adds to that specification an interaction between any publication and one worker characteristic, together with its main effect; the plotted coefficient is the interaction term. Experience is years since the Ph.D. divided by 10. Top Ph.D. equals 1 for the top eight programs (CMU, MIT, Berkeley, Stanford, Princeton, Cornell, Georgia Tech, and UIUC). BA in US/Asia indicate whether the worker obtained her bachelor's degree in the U.S. or Asia, respectively. Gender is identified from name and profile picture. First author indicates being listed first on the paper.

The first characteristic, labor market experience, proxies for belief precision: as opportunities to publish accrue over time, employers form a more precise belief about a worker's research ability. For both outcomes, more experienced workers show a smaller increase in upward mobility one year after publishing. Among workers more than ten years past the Ph.D., publishing has little effect, consistent with evidence that employer learning matters more early in the career (e.g., Lange 2007).

I consider a few other characteristics that may reflect varying degrees of precision of the prior beliefs about worker ability. The labor market may hold more precise priors about graduates from the top eight CS Ph.D. programs in the U.S. Any post-Ph.D. publication

raises transitions to top firms less for these graduates than for others, while its effect on becoming a research scientist at top firms does not differ. Workers with a bachelor's degree from Asia see a stronger increase in mobility to top firms after publishing, consistent with less precise priors about workers educated abroad, but this gap does not extend to becoming a research scientist at a top firm. The estimates for the gender interaction are noisy and centered around zero, providing no evidence for statistical discrimination in the form of less precise beliefs about either gender.

4.2.2 First Author versus Others

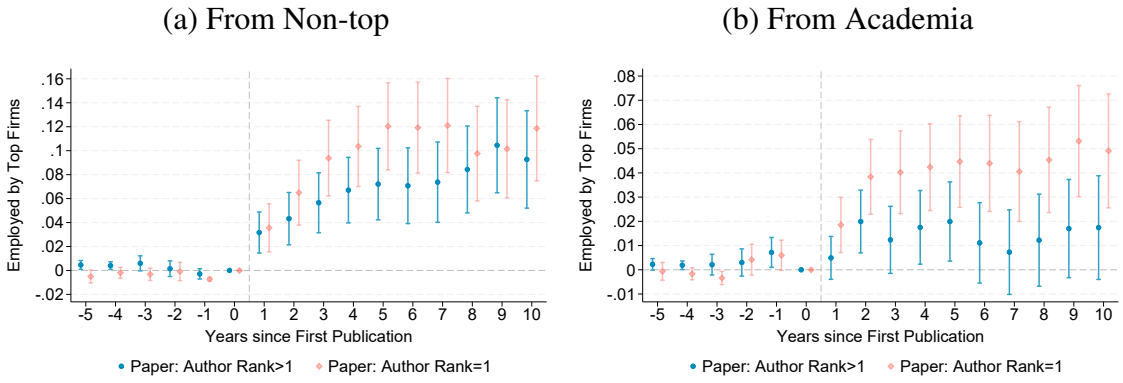
There are fewer opportunities to be the first author on a paper than to publish at all. I exploit that scarcity to test Prediction 3: fewer opportunities to produce a signal generate a larger belief update, and if the offer arrival rate is locally linear, a larger mobility response. The last row in Figure 3 shows that for workers at non-top firms, first authorship raises upward mobility to top firms by 1.1 p.p. ($t = 1.5$), and the probability of becoming a research scientist at a top firm by 1.7 p.p. ($t = 3.0$). The effect of first authorship is thus concentrated in the research-scientist outcome.

Focusing on the first publication one to three years after the Ph.D., I estimate a dynamic difference-in-differences regression that allows the dynamic effects to differ between first authors and other authors:

$$M_{it} = \sum_{l \neq 0} \beta_l D_{it}^{(l)} + \sum_{l \neq 0} \gamma_l^{\text{first}} D_{it}^{(l)} \cdot \text{Published}_i \cdot \text{FirstAuthor}_i \tag{9}$$
$$+ \sum_{l \neq 0} \gamma_l^{\text{others}} D_{it}^{(l)} \cdot \text{Published}_i \cdot (1 - \text{FirstAuthor}_i) + \alpha_i + \mu_t + \varepsilon_{it}.$$

Panel (a) of Figure 4 shows that first authors from non-top firms move up at similar rates to other authors in the first year after publication, but follow a steeper trajectory over years two to five. Panel (b) shows the same event study for workers whose first publication is in academia. There, first authors move to top firms at higher rates than other authors even in the first year, and the gap persists over the following decade. The sharper contrast in academia may reflect that author order is more informative about individual contributions there: industry employers may set author order by internal hierarchy or even conceal talent strategically, as in Waldman (1984), weakening the signal content of first authorship.

FIGURE 4. Event Study: Moving to Top Firms, First Author vs. Others



Notes: This figure plots the difference-in-differences estimates of $\{\gamma_l^{\text{first}}\}$ for first authors, and $\{\gamma_l^{\text{others}}\}$ for other authors, from equation (9). Panel (a) focuses on workers employed by non-top firms at event time $l = 0$, and panel (b) on workers in academia at $l = 0$. The treated publish their first paper 1–3 years after the Ph.D. The controls are assigned a placebo event time from the ordered logit described in Footnote 23. All workers have a job record in every year between -1 and 3 relative to the event time. Standard errors are robust and clustered at the worker level.

4.3 Additional Signals: Patent Applications matched to Papers

Firms often file patent applications to protect the inventions disclosed in a publication. Papers with a matched patent application show a steeper rise in citations after the first year (Appendix Figure OC.3), suggesting that the authors' employers observe more about a work's quality and commercial value than is disclosed in the paper, and file for a patent to secure exclusive rights.

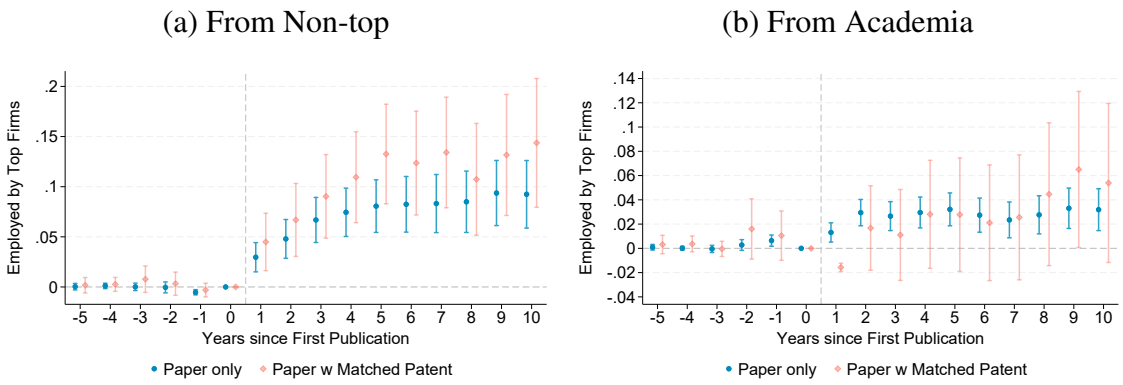
I identify patent applications matched to CS papers by comparing team members, affiliations, and content similarity.[25] About 25% of papers by CS Ph.D.s in industry, and 5% of papers by those in academia, are accompanied by a patent application. Appendix Table OA.3 provides examples: a matched pair may have different titles but similar abstracts, and the patent application often adds technical details and clarifies which contributions of the paper are claimed as inventions.

Since there are fewer opportunities to produce papers with a matched patent, I test Prediction 3 by replacing FirstAuthor_i in regression (9) with an indicator for whether the first publication has a matched patent application. Panel (a) of Figure 5 shows the estimates

[25] I follow the prediction-based matching of Marx and Scharfmann (2024); the matching procedure is described in Appendix C.2.

for workers at non-top firms: those whose first publication has a matched patent are 1.5 p.p. more likely to move to a top firm within a year than coworkers who publish without one, and the gap widens to 5.2 p.p. by year five, though it is never statistically significant at the 5% level.

FIGURE 5. Event Study: Moving to Top Firms, Matched Patent or Not



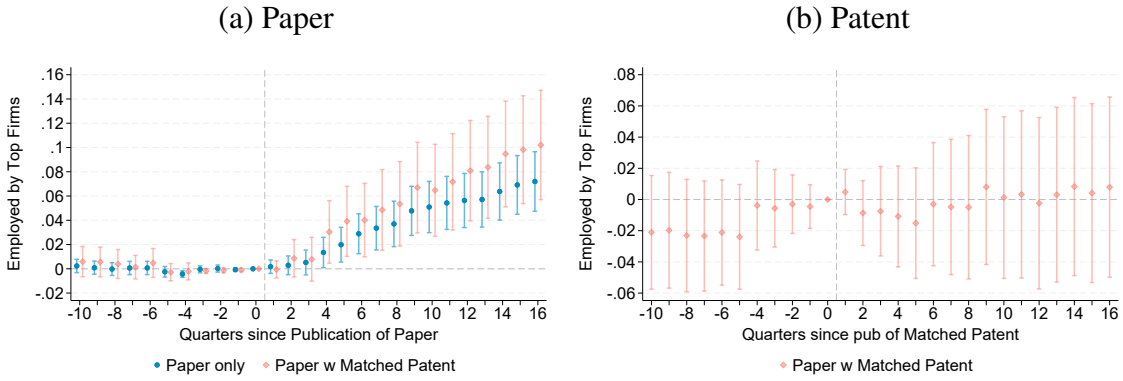
Notes: This figure shows the difference-in-differences estimates from equation (9), replacing the first-author indicator with an indicator for whether the first publication has a matched patent application (MP_i). I plot the dynamic effects $\{\gamma_l^{\text{MP}}\}$ for workers with $MP_i = 1$ and $\{\gamma_l^{\text{others}}\}$ for other authors. Panel (a) covers workers employed by non-top firms at $l = 0$, and panel (b) on workers in academia at 0. All workers are required to have a job record in every year between -1 and 3 relative to the event time. Standard errors are robust and clustered at the worker level.

Researchers in academia provide a placebo test. Academia does not hire or promote researchers based on patents, and academics have more discretion over whether to patent their work. The number of matched patents is therefore not an employer-set opportunity constraint N in academia, and the scarcity mechanism in Prediction 3 does not apply. Panel (b) shows the estimates for workers in academia at $l = 0$: a matched patent does not predict higher transitions to top firms, suggesting that top firms do not place much weight on patenting when screening candidates from academia.

Beyond serving as a rarer signal, the matched patent also arrives with a lag: patent applications are disclosed by the USPTO with an 18-month delay by default, so for a window the incumbent employer — who makes the patenting decision — knows of the patent application while outside firms do not. I find no evidence that workers advertise unpublished patent applications on LinkedIn, and conversations with industry professionals indicate that recruiters respect non-disclosure agreements and avoid asking about unpublished applications. I exploit this lag to test for asymmetric learning.

Focusing on the first publication within three years of the Ph.D., I re-estimate regression (9) at the quarterly level around the conference date, when the paper becomes public. Panel (a) of Figure 6 shows that among workers at non-top firms, those with a matched patent are over 1 p.p. more likely to be at a top firm four quarters after the paper date than authors without one, and the gap persists thereafter. As in the annual event study (panel (a) of Figure 5), the gap is not significant.

FIGURE 6. (Quarterly) Event Study around Paper versus Matched Patent



Notes: Panel (a) shows the difference-in-differences estimates of regression (9), replacing the first-author indicator with an indicator for whether a worker's first publication 1-3 years after Ph.D. has a matched patent application. Event time 0 is the quarter the paper is published (using conference dates). The controls are workers without a publication 1–3 years after the Ph.D. and I assign a placebo event quarter. Panel (b) restricts to workers who publish within 3 years, and the event time 0 is the quarter a matched patent is published by the patent office. The controls for (b) are workers whose first publication is not matched to any patent application. For the controls, I draw a placebo quarter of patent revelation by estimating a Poisson regression of quarters until patent disclosure on initial worker and job characteristics.

Panel (b) instead centers on the quarter the patent application is disclosed by USPTO. I focus on workers who publish 1-3 years after Ph.D., so the comparison is between matched-patent and paper-only authors. I assign a placebo event quarter to each worker in the paper-only group.[26] The difference-in-differences estimates show no increase in upward mobility to top firms around the patent disclosure date. The absence of a jump makes it hard to pin down when the information gap closes — it does not appear to close when the patent

[26] Among workers with a patent matched to their first publication, I estimate a Poisson regression of the number of quarters between paper and patent disclosure on characteristics of the first job and Ph.D. school×cohort fixed effects. For each worker in the paper-only group, I draw a placebo event time from a Poisson distribution with mean equal to the predicted value.

application becomes public.

Looking beyond the first publication after the Ph.D., I find some evidence of asymmetric learning based on the relationship between lagged innovation outputs and employment outcomes. At non-top firms, workers who publish a paper with a matched patent are 3.5 p.p. less likely to leave in one year than paper-only coworkers (column 1 of Table 1). Incumbent employers may match outside offers to retain workers they privately know to be better than the market believes.

TABLE 1. Papers & Matched Patents on Job Mobility

	Move to a different firm at (t+1)			Employed by Top Firms at (t+1)		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
Research Output at t						
Paper+Patent	0.0088	-0.0172**	0.0043	0.0133**	0.0136**	-0.0007
	(0.0108)	(0.0073)	(0.0082)	(0.0067)	(0.0062)	(0.0028)
Paper only	0.0446***	0.0039	0.0100***	0.0213***	-0.0000	0.0034**
	(0.0076)	(0.0052)	(0.0035)	(0.0043)	(0.0044)	(0.0014)
Research Output at $[t-3, t-1]$						
Paper+Patent	0.0236***	0.0102*	0.0041	0.0103**	-0.0044	0.0059**
	(0.0074)	(0.0056)	(0.0063)	(0.0046)	(0.0051)	(0.0026)
Paper only	0.0044	0.0055	0.0134***	0.0035	-0.0041	0.0053***
	(0.0042)	(0.0036)	(0.0032)	(0.0022)	(0.0033)	(0.0012)
Mean	0.1223	0.0690	0.1038	0.0251	0.9479	0.0090
N	1.34e+05	5.21e+04	7.23e+04	1.34e+05	5.21e+04	7.23e+04
Adjusted R2	0.1126	0.0253	0.1139	0.0307	0.0190	-0.0340

Notes: This table presents regressions of employment outcomes at $t+1$ on indicators for research output at t, and at $[t-3, t-1]$. The sample is at the person×year level, including years with full-time employment after the Ph.D. "Paper only" is an indicator for a worker having any paper but no matched patent. "Paper+Patent" is an indicator for publishing a paper with a matched patent. The $[t-3, t-1]$ block defines the analogous indicators over the prior three years. Columns (1)-(3) use a move to a different firm as the dependent variable. Columns (4)-(6) use employment by a top firm at $t+1$. I estimate separate regressions for workers currently at non-top firms, top firms, or academia. All regressions use the same controls and fixed effects described under Figure 1. Standard errors are robust and clustered at the Ph.D. school×cohort level.

In contrast, a matched patent filed in the past three years is more likely to be public. Workers with a matched patent during that period are 2.4 p.p. more likely to move in the next year (column 1), and 1.0 p.p. more likely to enter a top firm (column 4), relative to coworkers without a paper; both gaps are statistically significant. The gaps between

authors with papers alone in $[t-3, t-1]$ and coworkers without papers are closer to zero, conditional on controls for research output at t. These patterns echo the higher upward mobility among authors with a matched patent in Figures 5(a) and 6(a). But the absence of a jump at patent disclosure in Figure 6(b) casts doubt on whether the information asymmetry is truly about the matched patent, or more generally reflects that outside employers take more time to recognize higher-quality research.

In summary, Section 4 emphasizes the role of publications after the Ph.D. in predicting J2J transitions and upward mobility to top firms, especially for workers at non-top firms. These patterns provide evidence for employer learning from public information — conference publications — consistent with the predictions from the benchmark model that focuses on symmetric learning in an equilibrium search framework. The evidence for asymmetric learning is less clear, and hence I focus on public learning in the structural analysis.

5 Structural Analysis of the Impact of Learning on Productivity

The empirical analysis so far highlights the differential impact of publications on job mobility for workers in academia, non-top industry firms, and top firms. Publishing today also strongly predicts future output — authors produce about 5 times as many publications over the next five years as coworkers without a publication, across origins (Figure OA.6) — but this reduced-form persistence cannot separate employer learning from fixed ability differences.

I present a model extension that distinguishes between groups of employers and allows transitions across groups to depend flexibly on market belief. I estimate the model and assess how much employer learning from publications contributes to sorting of workers across firms, and in turn to aggregate publication productivity.

5.1 Model Extension

To take the model to the data, I introduce a segmented labor market with three groups of employers: academia (A), non-top industry firms (I), and top industry firms (T). The three groups provide different opportunities to publish, so aggregate publication output depends on labor market sorting. Within each group, employers are sorted by productivity

into rungs of a job ladder as in the benchmark model. In addition, cross-group shocks reallocate workers across (A, I, T), depending on employer belief and experience.

Separating academia from industry is natural: workers in academia move less frequently and publish more (Table OA.2). The split between non-top and top firms in industry is less standard but is justified by two empirical patterns. First, ranking industry employers by revealed preference or average wages, I find that workers who publish at non-top firms are more likely to move to a top firm but not to higher-ranked non-top firms within I. Second, academics who publish are more likely to move to top firms but not to other firms in industry. Both patterns suggest that top firms occupy a distinct position in hiring from both industry and academia: were top firms simply the highest rung of a single industry ladder, publications would predict upward mobility throughout the ladder rather than to T alone.

I assume employers observe workers' initial characteristics X, such as educational background and pre-Ph.D. publications, as well as additional signals that may not be observed by econometricians but are important for the first placement after the Ph.D. For labor market entrants with characteristics X starting in group $k \in \{A, I, T\}$, employers share a prior $\theta \sim \text{Beta}(\alpha(X,k), \beta(X,k))$, in which the shape parameters depend on both X and the first employer group.

Given the heterogeneity in the opportunities to publish across groups, employer learning now depends on a worker's job history. Letting N_{kt} denote the number of i.i.d. Bernoulli trials for workers in group k at experience t, the number of publications conditional on ability θ is distributed according to $\text{Binomial}(N_{kt}, \theta)$. Suppose the current employer belief is $\text{Beta}(\alpha, \beta)$. The posterior becomes $\text{Beta}(\alpha + y, \beta + N_{kt} - y)$ for a worker who produces y publications while employed in group k at experience t.[27]

Denote by $\Lambda_{kk'}(E[\theta|\alpha, \beta], t)$ the probability with which a worker with employer belief $\text{Beta}(\alpha, \beta)$ and employed in group k receives a job offer from group k'. The diagonal terms Λ_{kk} play the role of λ_1 in the benchmark model: they are within-group offer arrival rates, and the worker only accepts offers that yield a higher joint value than her current match. The off-diagonal terms $\Lambda_{kk'}$ for $k \neq k'$ are arrival rates of cross-group offers. I assume workers cannot bargain across groups, so the cross-group offers are "Godfather" shocks that workers do not refuse (e.g., Jolivet, Postel-Vinay, and Robin 2006).[28] A worker

[27] Because the prior varies with (X, k) and opportunities vary across groups, cumulative publications and experience are no longer sufficient statistics for the employer belief; the belief parameters (α, β) themselves serve as the public state.

[28] For workers who receive a cross-group offer, I assume the fallback option is unemployment and therefore the offer from a different group is always accepted. Without this assumption, comparing an offer from

exposed to such a shock draws her new employer in k' from the offer distribution $F_{k'}$.

I also denote by Λ_{0k} the rate at which an unemployed worker receives an offer from group k, the analog of λ_0 in the benchmark model. Workers are laid off at a rate δ common to all groups, while the within-group involuntary J2J rate κ_k is group-specific.

To keep the state space finite, I cap experience at $t = 10$: beyond that point, both the employer belief and the experience level remain fixed. This treats $t = 10$ as an absorbing state and is motivated by the empirical finding that employer learning matters most for less experienced workers (e.g., Altonji and Pierret 2001; Lange 2007). The steady-state flow-balance equations extend those in the benchmark model (equations (4)–(5)) by tracking inflows and outflows across the three groups. Full details of the model extension are in Appendix D.

5.2 Identification and Estimation

I estimate the model by the simulated method of moments (SMM). I discuss the model parameters (Appendix Tables A.2-A.3) and the moments that identify them (Appendix Table A.1) in detail below.

5.2.1 Job Ladders

Following Bagger and Lentz (2019) and Sorkin (2018), I rank firms by their empirical job-to-job (J2J) transitions using the PageRank algorithm (Page, Brin, Motwani, and Winograd 1999), which assigns a higher value to firms that receive a greater net inflow from other firms. This procedure has a revealed-preference interpretation consistent with the model: under efficient contracting, workers accept within-group outside offers only when the new match yields a higher joint value, and the joint value is strictly increasing in firm productivity (Lemma 1).[29] Net J2J flows therefore reveal the productivity ranking, with PageRank summarizing these flows into firm-level values.

I run the PageRank algorithm separately for industry and for academia, excluding firms outside the largest connected component. The highest-valued industry employers are

industry with a current job in academia, for example, would require modeling non-wage amenities and idiosyncratic preferences.

[29] The flow output (1) rises with firm productivity ϕ, while the gain from poaching declines with ϕ. Lemma 1 shows that the former force dominates, because the mutually exclusive shock probabilities sum to less than one. The argument extends to each employer group in the extended model, whose joint value is presented in equation (A.16) in Appendix D.

Alphabet, Facebook, Amazon, Apple, and Microsoft, all of which are labeled as "top firms" in the analysis so far.[30] IBM ranks slightly lower but remains one of the largest employers of CS Ph.D.s and invests heavily in publishing and patenting. I place these six firms in T, assuming that they have the same productivity and opportunities to publish. I then partition non-top firms into six rungs in I based on their PageRank values. The rungs serve as a discrete approximation to the continuous productivity distribution in the model.

The top 20 academic employers according to PageRank comprise 12 U.S. public universities, 5 U.S. private universities, and 3 universities overseas. Given the distinct mobility patterns of postdocs relative to tenure-track and tenured positions, I use job titles to place all postdocs and non-tenure-track researchers at the bottom rung of the academic ladder.[31] I sort the remaining academic positions into three rungs based on PageRank values.

5.2.2 Employer Beliefs and Publication Opportunities

Workers are endowed with research ability $\theta \in [0,1]$ drawn from $\text{Beta}(\alpha(X,k),\beta(X,k))$ given initial characteristics X and the group k of their first job at $t=0$. The vector X includes the ranking of a worker's Ph.D. institution, pre-Ph.D. publications, and demographics. The model does not explicitly address how new Ph.D.s are placed into their first jobs. I take the first employer as given, and parameterize the shape parameters in the prior as follows:

$$\alpha(X,k) = \exp(a'X + \alpha_k),\ \ \beta(X,k) = \exp(b'X + \beta_k) \tag{10}$$

where a and b are coefficient vectors on the observables X, while α_k, β_k are common among workers initially placed in group $k \in \{A, I, T\}$. Employers may observe additional information about new Ph.D.s that drives the initial sorting. This parameterization allows such unobserved heterogeneity, but restricts it to be common at the employer-group level. The combination of a higher α_A and a lower β_A in academia than in other groups would suggest that workers placed in academia are higher-ability on average.

Given the opportunities to publish in each group at $t=0$, (N_{A0}, N_{I0}, N_{T0}), the parameters in (10) are identified by matching the first and second moments of the number of publications at $t=0$ separately by group, and the regression coefficients of publication

[30] The top five remain the same if I use PageRank to provide a single ranking over academic and industry employers. Figure OA.7 shows a positive relationship between the value returned by PageRank and mean wages according to H-1B records.

[31] For example, a postdoc at Carnegie Mellon University (CMU) is assigned to the lowest rung of the academic ladder despite CMU's high PageRank.

output (quantity and indicator) on X and group dummies (see Step 1 in Table A.1). In an outer loop, I consider 285 possible (N_{A0}, N_{I0}, N_{T0}), each of which has $N_{k0} \in \{2, \ldots, 10\}$ and $N_{I0}, N_{T0} \leq N_{A0}$.[32]

For $t \geq 0$, the opportunities to publish in each group are parameterized as:

$$N_{kt} = \left[N_{k0} \times \exp(\nu_{k1} t + \nu_{k2} t^2 + \nu_{k3} t^3) \right] \tag{11}$$

where $[\cdot]$ denotes rounding to the nearest integer. The unrestricted coefficients in (11) are identified from the group-specific means of publications across experience levels and the regression coefficients of publications on a polynomial of experience.[33]

5.2.3 Labor Market Transitions

Given employer belief Beta(α, β), cross-group transition rates $\Lambda_{kk'}$ are parameterized as functions of the posterior mean $\pi := E[\theta|\alpha,\beta] = \frac{\alpha}{\alpha+\beta}$ and experience t:

$$\Lambda_{kk'}(\pi, t) = r_{kk',t} \cdot \Lambda_{kk',0} \cdot \exp(\Lambda_{kk',1} \pi). \tag{12}$$

The sequence $\{r_{kk',t}\}$ is set to the empirical decay profile of cross-group transition rates with experience. In the exponential form above, $\Lambda_{kk',1}$ is the semi-elasticity of the cross-group transition rate with respect to the posterior mean π. The specification can be viewed as the reduced form of a discrete-choice model in which workers draw idiosyncratic preferences over groups from a Type-I Extreme Value distribution, as in Card, Cardoso, Heining, and Kline (2018) and Lentz et al. (2026): when the transition probabilities are small, $\Lambda_{kk',1}$ approximates the coefficient on π in the utility of choosing k' relative to the origin k.[34] The semi-elasticities $\Lambda_{kk',1}$ are identified by the empirical relationships between transition rates and publication output, measured either as the inverse hyperbolic sine (asinh) of publications or as a binary indicator for any publication. The baseline constants $\Lambda_{kk',0}$ are directly identified by the average cross-group transition rates.[35]

[32] The restriction that there are at least as many opportunities to publish in academia as in either industry group trims the grid search. It does not drive the estimates, as the selected triple $(4,2,3)$ is interior to the restriction. Ability differences across groups are separately captured by group-specific priors as in equation (10), which are identified jointly with N_{k0}'s from the first and second moments of publications at $t = 0$.

[33] I fit a cubic polynomial for T, a quadratic for A ($\nu_{A3} = 0$), and a linear profile for I ($\nu_{I2} = \nu_{I3} = 0$). All unrestricted coefficients are estimated (Table A.3).

[34] Suppose the deterministic utility of choosing k' over the origin k is $u(\pi) = a + b\,\pi$, normalizing the utility of staying in k to zero. Under Type-I EV idiosyncratic preferences, the choice probability of moving from k to k' is $p(\pi) = \exp(u(\pi))/(1+\exp(u(\pi)))$. The semi-elasticity $d\log p/d\pi = b(1-p) \approx b$ when p is small.

[35] As shown in Table A.3, $\Lambda_{IA,0}$ (industry to academia) and $\Lambda_{AI,0}, \Lambda_{AT,0}$ (academia to non-top and top industry firms, for tenure-track/tenured workers) are set to the empirical means because these transition rates

The within-group offer arrival rates Λ_{AA} and Λ_{II} are also parameterized as in (12).[36] The baseline constants $\Lambda_{kk,0}$ are identified from origin-specific transition rates to higher-ranked rungs within $k \in \{A, I\}$. The semi-elasticities $\Lambda_{kk,1}$ are identified by the empirical relationship between within-group upward mobility and publication output.

Within academia, postdocs (at the lowest rung of A) have higher J2J rates than tenure-track workers. For each destination $k \in \{A, I, T\}$, I allow a postdoc premium $\Lambda^p_{Ak,0}$ on the baseline constant and a separate semi-elasticity $\Lambda^p_{Ak,1}$, so that the arrival rate for postdocs is $\Lambda^p_{Ak}(\pi, t) = r_{Ak,t} \cdot (\Lambda_{Ak,0} + \Lambda^p_{Ak,0}) \cdot \exp(\Lambda^p_{Ak,1}\pi)$. The postdoc-specific parameters are identified from origin-specific J2J rates and from the interaction of publications with postdoc status in mobility regressions (Table A.1).[37]

The offer distributions F_I and F_A are parameterized as $F_k(j) = \frac{\sum_{j' \leq j} \exp(v_k(j'))}{\sum_{j''} \exp(v_k(j''))}$, where j indexes the rungs in group k and $v_k(0)$ is normalized to 0. The remaining $v_k(j)$ are identified from origin-specific upward and downward J2J moves within each group.

The remaining transition probabilities do not vary with the employer belief and are estimated directly from the data (Table A.3). The layoff rate δ is identified by the empirical employment-to-unemployment (E2U) rate. The involuntary J2J rate κ_k is identified by the rate of involuntary downward J2J moves within each group.[38] The rates at which the unemployed receive offers from each group, $\{\Lambda_{0k}\}$, are identified by the empirical unemployment-to-employment (U2E) rates. The labor market exit rates, μ for $t < 10$ and $\overline{\mu}$ at $t = 10$, are identified by the rates at which workers leave the labor force.

5.2.4 Estimation Procedure

The model is estimated by the simulated method of moments (SMM). I randomly assign 75% of computer scientists in the data to a training set and the remainder to a holdout sample for validation.[39] There are 42 continuous parameters, along with the integer triple

are close to zero in the data and difficult to estimate reliably given the boundary constraint at zero. I assume $\Lambda_{IA,0} = \Lambda_{TA,0}$ and $\Lambda_{IA,1} = \Lambda_{TA,1}$.

[36] Λ_{TT} is dropped given the assumption that top firms in T share the same productivity ϕ.

[37] $\Lambda_{AA,0}$ (within-academia upward J2J for tenure-track/tenured workers) is set to the empirical mean, which allows the postdoc premium $\Lambda^p_{AA,0}$ to be identified separately via SMM.

[38] Since top firms share the same productivity, within-T reshuffling is irrelevant, and I set $\kappa_T = 0$. Involuntary moves from T to I are instead captured by a cross-group constant Λ_{TI}, identified by the empirical T-to-I transition rate.

[39] The estimation sample contains 26,528 computer scientists who graduated between 2000 and 2021, have a job record in the year after the Ph.D., and are ever employed by a firm in the largest connected component. Table OA.1 shows that the estimation sample is similar on average to the full sample of graduates in the same window.

of initial publication opportunities $\vec{N}_0 \coloneqq (N_{A0}, N_{I0}, N_{T0})$, which is selected by grid search as described below. The estimator minimizes the weighted distance between the empirical moments, computed from the training set, and their simulated counterparts.[40]

The model parameters in Table A.1 are estimated in two steps. First, I focus on $\vec{N}_0$ and the 14 parameters that enter the prior in equation (10). For each candidate $\vec{N}_0$, I simulate labor market entrants by bootstrapping workers at $t = 0$ with replacement from the training data, taking their initial employer and observables X as given. I draw true ability θ from $\text{Beta}(\alpha(X,k), \beta(X,k))$ given a guess of the prior parameters, and publication output from $\text{Binomial}(N_{k0}, \theta)$ for workers who start in group k. The moments for this step concern publication outputs at $t = 0$ and their relationship with the observables X and group k (Step 1 in Table A.1). I screen all 285 candidate triples from a single initial guess of the prior parameters, keep the best-fitting triples, and re-estimate the prior parameters under each $\vec{N}_0$ from multiple initial guesses; the best combinations are carried to the next step.

The second step estimates the remaining 28 parameters — the experience profiles of publication opportunities in (11), the transition rates, and the offer distributions — from publication outputs and labor market transitions in the full panel, where experience $t \in \{0, \ldots, 10\}$ (Step 2 in Table A.1). Given parameters from the first step and a guess of the remaining parameters, I simulate the full panel by initializing entrant cohorts as in the first step and advancing incumbent workers through the timing of the model: the mutually exclusive shocks are realized — labor market exit, layoff, and advance layoff notice at exogenous rates, and outside offers arriving at rates that depend on the current posterior mean; workers transition to new employers as applicable; publication output is drawn from $\text{Binomial}(N_{kt}, \theta)$; and employer beliefs are updated via conjugate Beta-Binomial updating. I simulate 50 periods and drop the first 25 as burn-in, so that the simulated panel reaches its stationary distribution. Further details are in Appendix E.

In both steps, I minimize the weighted distance between the simulated and empirical moments via the L-BFGS-B algorithm, a quasi-Newton technique that supports box constraints on parameters (Byrd, Lu, Nocedal, and Zhu 1995; Zhu, Byrd, Lu, and Nocedal 1997). All simulation draws are held fixed across parameter evaluations, so that changes in the objective reflect the parameters rather than simulation noise. To reduce the risk of converging to a local minimum, I start the optimization from multiple initial guesses in both steps, refine the search from the best-performing parameter vectors, and report the estimates that achieve the best fit in the holdout sample (Tables A.2-A.3).

[40] See equation (A.21) in Appendix E for the shrinkage weight matrix.

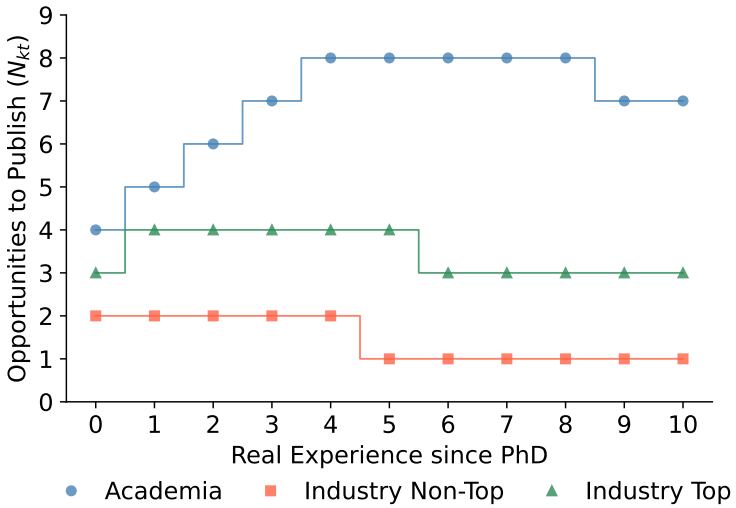
5.3 Model Fit

I assess how well the estimated model reproduces the empirical moments, focusing on (1) publication output and the resulting belief updates, (2) transitions to top firms, and (3) transitions between academia and industry. The model matches the publication moments closely and reproduces the empirical mobility responses to publication signals.

5.3.1 Publication Outputs and Employer Beliefs

The first set of moments concerns publication output across employer groups and experience. The grid search in Step 1 estimates 4 opportunities to publish for entrants starting in academia, 3 for those starting at top firms, and 2 for those starting at non-top firms (Figure 7). Almost all simulated publication moments at $t = 0$ fall within the 95% CI of the empirical moments in the training set; the exceptions are the share with any publication in academia and the second moment of publications in industry (Figure OA.8). Table A.2 reports the estimated prior parameters in (10). Entrants with more pre-Ph.D. publications have higher prior means, as do entrants starting in academia or top firms relative to non-top firms; other characteristics show no significant impact. The coefficients entering the shape parameter β are weakly identified, which accounts for the missed moments above.

FIGURE 7. Opportunities to Publish

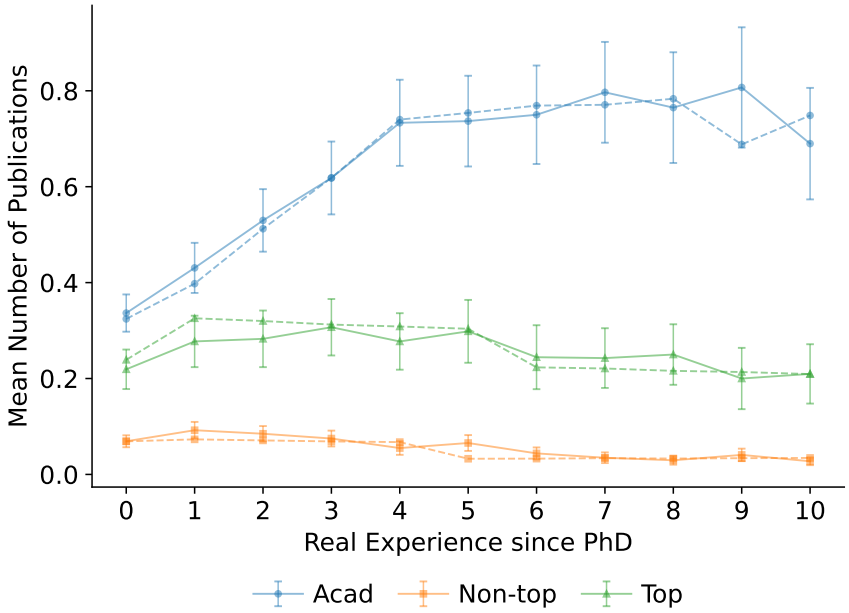


Notes: This figure shows the estimated number of opportunities to publish (Bernoulli trials) for workers in each group k. The initial $\{N_{k0}\}$ are estimated via a grid search in Step 1, and the changes with experience parameterized in (11) are estimated in Step 2 (see Table A.1).

Based on the estimates of equation (11) in Step 2, the opportunities to publish in academia, N_{At}, continue to rise and stabilize at 8 per year between $t = 4$ and $t = 8$, falling slightly to 7 at $t \geq 9$ (Figure 7). At top firms, N_{Tt} increases from 3 to 4 in the first year and drops back to 3 after five years. At non-top firms, N_{It} never rises and equals 1 at $t \geq 5$. The gap between academia and industry, and between top and non-top firms within industry, therefore widens over the career.

Figure 8 shows that the estimated model closely traces the experience profiles of mean publication output by employer group, measured in the holdout sample. Most simulated moments fall within the 95% bootstrapped CI of the empirical moments. Consistent with the greater opportunities to publish in academia, the research productivity gap between academics and workers in industry widens over the first five years of their careers. Those at top firms are more likely to publish than those at non-top firms, and the gap between them shrinks only modestly after $t = 6$.

FIGURE 8. Publication Output by Employer Group and Experience



Notes: This figure shows the empirical (solid) versus simulated (dashed) mean number of publications by real experience t, separately by employer group. The empirical moments are measured in the 25% holdout sample, which does not enter the SMM objective. Error bars are 95% bootstrapped confidence intervals around the empirical moments.

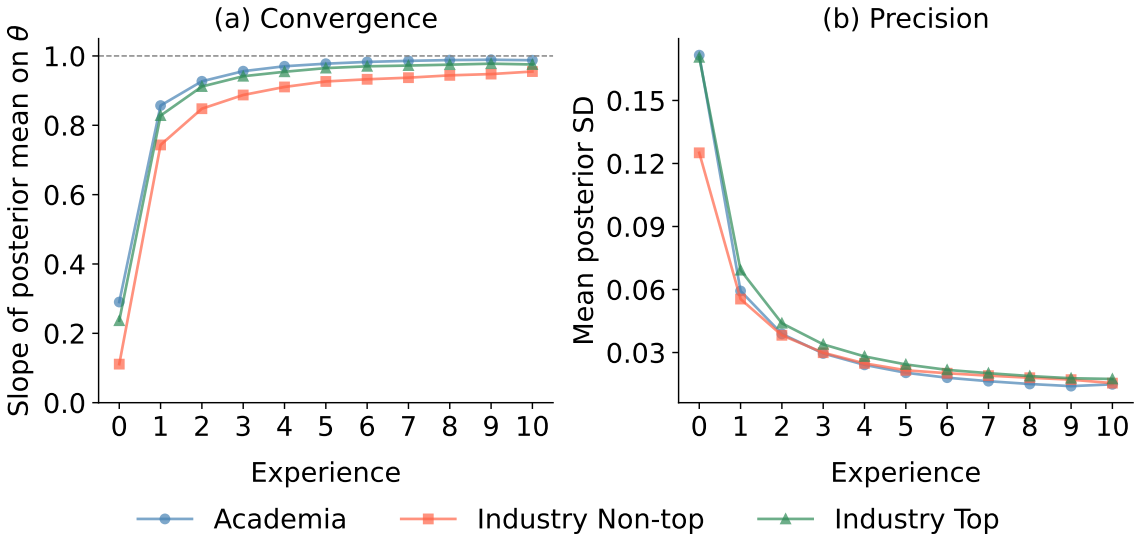
The priors based on initial placement and characteristics increase with true ability θ,

as shown in the binned scatter plot for $t = 0$ in Figure OA.9. Once the labor market observes on-the-job publications, the posterior means become more aligned with true ability, approaching the 45-degree line after five years of experience.

Following Lange (2007), I also examine the speed of employer learning, which depends on the number of opportunities to publish under the Beta-Binomial update scheme. Panel (a) of Figure 9 shows the regression coefficient of the posterior mean on true ability, separately by experience and employer group. The coefficient converges towards one in all three groups, consistent with Figure OA.9. The convergence is faster for workers in academia and top firms, which provide more opportunities to publish than non-top firms (Figure 7).

Panel (b) of Figure 9 shows that employer beliefs become more precise over time, as the mean standard deviation of beliefs falls. At $t = 0$, uncertainty is greater for workers in academia or top firms. Precision improves quickly once the labor market observes post-Ph.D. publications, and the improvement is largest in academia, which provides the most opportunities to publish. The standard deviations converge across groups after $t = 5$.

FIGURE 9. Belief Update by Employer Group



Notes: This figure shows the convergence of employer beliefs towards true ability and the precision of beliefs over experience. Given belief Beta(α, β), the posterior mean is $\frac{\alpha}{\alpha+\beta}$ and the standard deviation is $\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$. Panel (a) plots the coefficient from regressing the posterior mean on true ability θ in the simulated panel, separately by experience and employer group. Panel (b) plots the mean standard deviation of beliefs.

5.3.2 Job-to-Job Transitions to Top Firms

The estimated semi-elasticity of moving from non-top to top firms with respect to the posterior mean is 1.36 (s.e. 0.40), indicating that workers with higher employer beliefs sort into top firms ($\Lambda_{IT,1}$ in Table A.3). A key moment for identifying $\Lambda_{IT,1}$ is the regression coefficient of mobility from non-top to top firms on publication output each period, measured by the asinh of publications.

TABLE 2. Transitions to Top Firms: Data versus Simulations

	From Industry-Nontop			From Academia		
	(1) Train	(2) Holdout	(3) Sim	(4) Train	(5) Holdout	(6) Sim
Asinh(Paper)	0.0323***	0.0314***	0.0325***	0.0249***	0.0162**	0.0333***
	(0.0054)	(0.0087)	(0.0018)	(0.0041)	(0.0068)	(0.0013)
Asinh(Paper) $\times$ t	-0.0019**	-0.0029**	-0.0011***			
	(0.0008)	(0.0013)	(0.0002)			
Asinh(Paper) $\times$ Tenure				-0.0249***	-0.0161**	-0.0325***
				(0.0042)	(0.0068)	(0.0013)
Experience	-0.0013***	-0.0010***	-0.0017***			
	(0.0001)	(0.0002)	(0.0000)			
Tenure-track				-0.0163***	-0.0191***	-0.0112***
				(0.0012)	(0.0023)	(0.0003)
Constant	0.0324***	0.0294***	0.0351***	0.0192***	0.0218***	0.0147***
	(0.0009)	(0.0016)	(0.0003)	(0.0012)	(0.0022)	(0.0003)
Observations	109,899	36,043	1,996,372	46,517	16,083	747,289
Adjusted R^2	0.0024	0.0015	0.0028	0.0133	0.0118	0.0144

Notes: This table shows the regressions of transitions to top firms (T) from non-top industry firms (I) in columns (1)–(3) and from academia (A) in columns (4)–(6). Columns (1) and (4) are estimated on the 75% training sample used to estimate the model; columns (2) and (5) on the 25% holdout sample; and columns (3) and (6) on the simulated data. Each simulation has 1,250 workers entering the market each period and runs for 50 periods. I drop the first 25 periods of each simulation and concatenate the simulated panel across 6 rounds with different seeds to reduce the impact of simulation noise. *p<0.1; **p<0.05; ***p<0.01.

The first three columns of Table 2 present this regression, estimated on the training data, the holdout sample, and the simulated data, respectively. The coefficient on As-

inh(Paper) in the simulated data matches the training coefficient and falls within the 95% CI of the holdout estimate. Comparing workers with 0 versus 1 publication, the estimates imply a roughly 3 p.p. increase in the probability of moving to top firms in the next year. Consistent with Figure 3, the marginal effect of publications decreases with experience.

I also estimate event-study regressions around the first publication, among workers employed at non-top firms when publishing their first paper after the Ph.D. The event-study coefficients are not targeted in the estimation and thus provide an out-of-sample check on the model. As shown in panel (a) of Figure OA.10, the share of authors employed by top firms reaches 6% in the first year after publication and 20% within five years, in both the data and the simulation. Following the difference-in-differences analysis in Figure 2, I assign a placebo event time to workers without publications and plot the estimates for employment at top firms of publishing authors relative to non-publishing workers. The model underestimates the immediate 2.5 p.p. increase by about 1 p.p., but the overall trajectories are similar: most simulated estimates are smaller than the empirical ones yet fall within their 95% CIs.

For postdocs in academia (the omitted category in columns 4-6 of Table 2), publications are also positively associated with transitions to top firms. The estimated model overestimates this relationship by 0.8 p.p. (column 6), though the simulated coefficient sits at the edge of the 95% CI of the training estimate (column 4). Tenure-track and tenured workers are 1.6–1.9 p.p. less likely than postdocs to move to top firms on average. The model underestimates this gap by 0.5 p.p., but matches the finding that publications have roughly zero impact on mobility to top firms for faculty.

5.3.3 Transitions between Industry and Academia

Workers who publish in industry are more likely to move to academia, and the effect is larger for the less experienced, as shown in columns (1)–(3) for those at non-top firms and columns (4)–(6) for those at top firms (Table 3). The 1.97 p.p. estimated effect of publications on transitions from non-top firms to academia in the simulated data (column 3) is very close to the 2.00 p.p. estimate in the training data (column 1) and falls within the 95% CI of the holdout estimate (column 2). For workers at top firms, the effect of publications on transitions to academia is smaller in each sample; the simulated estimate (column 6) is about one standard deviation above the training estimate and within one standard deviation of the holdout estimate.

TABLE 3. Transitions between Industry and Academia

	Non-top to Academia			Top to Academia			Academia to Non-top		
	(1) Train	(2) Holdout	(3) Sim	(4) Train	(5) Holdout	(6) Sim	(7) Train	(8) Holdout	(9) Sim
Asinh(Paper)	0.0200***	0.0342***	0.0197***	0.0141***	0.0220***	0.0178***	-0.0211***	-0.0014	-0.0030**
	(0.0040)	(0.0085)	(0.0012)	(0.0034)	(0.0069)	(0.0010)	(0.0049)	(0.0103)	(0.0015)
Asinh× t	-0.0018***	-0.0024**	-0.0014***	-0.0010**	-0.0020**	-0.0013***			
	(0.0006)	(0.0012)	(0.0001)	(0.0005)	(0.0009)	(0.0001)			
Asinh× tenure							0.0168***	-0.0005	0.0016
							(0.0049)	(0.0104)	(0.0015)
Experience	-0.0010***	-0.0010***	-0.0008***	-0.0007***	-0.0006***	-0.0008***			
	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0002)	(0.0000)			
Tenure-track							-0.0976***	-0.0998***	-0.1001***
							(0.0028)	(0.0048)	(0.0008)
Constant	0.0118***	0.0120***	0.0102***	0.0071***	0.0064***	0.0100***	0.1109***	0.1102***	0.1142***
	(0.0006)	(0.0010)	(0.0002)	(0.0009)	(0.0014)	(0.0003)	(0.0027)	(0.0047)	(0.0008)
Observations	109,899	36,043	1,996,372	35,535	11,470	804,882	46,517	16,083	747,289
Adjusted R^2	0.0036	0.0066	0.0031	0.0063	0.0103	0.0059	0.0506	0.0551	0.0506

Notes: Columns (1)–(3) show the regression estimates of transitions from non-top firms to academia on publication output; columns (4)–(6), transitions from top firms to academia; and columns (7)–(9), transitions from academia to non-top firms. The first column in each set (columns 1, 4, 7) is estimated on the 75% training sample used to estimate the model; the second column on the 25% holdout sample; and the last column on the simulated data, constructed as described in the notes to Table 2. *p<0.1; **p<0.05; ***p<0.01.

In contrast with the transitions to top firms in Table 2, academics who publish more are less likely to move to non-top firms in industry. The estimates on Asinh(Paper) for postdocs range from -2 p.p. in the training data to 0 p.p. in the holdout sample; the simulated estimate is negative and closer to the holdout estimate (column 8). For faculty, the relationship between publications and transitions to non-top firms remains negative across columns (7)–(9). The model matches the 10 p.p. baseline gap between postdocs and faculty in transitions to non-top firms.

Overall, the model reproduces the key features of the data: publication differences across academia, non-top firms, and top firms; the evolution of publications with experience; and the rise in transitions to top firms and to academia with publications. In Step 1, 14 of 17 simulated moments fall within the 95% bootstrapped CIs of the training moments (Figure OA.8), versus 12 holdout moments within the same intervals. In Step 2, 31 simulated moments fall within the 95% CIs of the 52 training moments, slightly fewer than the 35 holdout moments within the same intervals. At least some of the misses reflect sampling noise rather than misspecification. The model fits less well on rung-specific upward and downward moves among non-top firms. Given the goal of quantifying the impact of employer learning from publications on productivity, it is reassuring that the publication moments and the job-to-job transitions between academia, non-top firms, and top firms are well matched.

5.4 How Much Does Employer Learning Matter for Productivity?

Given the estimated model, overall publication productivity depends on how workers with varying ability sort among academia, non-top firms, and top firms over time. To quantify how much employer learning matters for productivity, I provide a Shapley-value decomposition of six mechanisms shaping sorting across employers:

1. Employer learning from publications after Ph.D.;
2. Initial sorting across academia, non-top firms, and top firms;
3. Inclusion of gender in priors;
4. Inclusion of whether the bachelor's degree is from the U.S. in priors;
5. Inclusion of Ph.D. institution ranking in priors;
6. Inclusion of pre-Ph.D. publications in priors.

The first mechanism is employer learning from on-the-job publications: when it is shut down, market beliefs remain fixed at the prior regardless of realized publications. The

second mechanism is initial sorting: when it is on, I take each worker's first job at $t = 0$ as given, thereby allowing employers to observe more about entrants than econometricians can see in the data — for example, interview signals that matter for the initial placement. When it is off, I randomly assign the first job, preserving the aggregate allocation of entrants across employers, and market priors are formed from initial observables alone.

Mechanisms 3–6 govern the information used to form priors. The estimated model uses priors based on all initial observables, as parameterized in equation (10); the distribution of true ability is held fixed in all counterfactuals.[41]

I estimate the average marginal impact of each mechanism on publication and sorting outcomes, following Shapley (1953).[42] I compute counterfactual outcomes under all $2^6 = 64$ subsets of mechanisms.[43] Letting $\hat{\Gamma}$ denote the estimated model parameters, the Shapley value of mechanism $m \in \mathbb{M} = \{1, \ldots, 6\}$ for outcome Z_{it} is:

$$SV_m = \sum_{C \subseteq \mathbb{M} \setminus \{m\}} \frac{|C|! \times (|\mathbb{M}| - |C| - 1)!}{|\mathbb{M}|!} \times \underbrace{\left(E\left[Z_{it} | C \cup \{m\}; \hat{\Gamma}\right] - E\left[Z_{it} | C; \hat{\Gamma}\right]\right)}_{\text{Change in mean outcome when adding mechanism } m} \tag{13}$$

The first column of Table 4 shows that allowing employers to update beliefs from on-the-job publications raises mean publications per worker-year by 0.03, a 14.28% gain relative to the benchmark with all mechanisms active. By employer group, learning raises publication output by 20% in academia, and by 17% at top firms, while lowering it by 32% at non-top firms — or 337 more publications per year in academia and 103 more at top firms, which together more than offset the 87 fewer at non-top firms.

The overall gain is driven by stronger sorting of higher-ability workers into top firms and academia, where they publish more than at non-top firms (Figure 8). For workers in the top 15% of the ability distribution, employer learning increases employment at top firms by 12.49% ten years after the Ph.D., and by 19.96% among those who start at non-top firms. The academic employment of this higher-ability group also increases by 18.60%. Meanwhile, the bottom 85% become less likely to end up at top firms (−3.29%) or in

[41] Appendix E explains how counterfactual priors are constructed when an observable is omitted. In a counterfactual without pre-Ph.D. publications, for example, the market prior is replaced by the Beta distribution matching the mean and variance of ability conditional on the remaining observables — i.e., integrating over the omitted characteristic. The prior mean is preserved, while the prior precision falls by the ability dispersion that the omitted characteristic accounted for.

[42] Shapley values have been applied to attribute model prediction or goodness of fit to individual features (e.g., Grömping 2007, Lindeman and Gold 1980). Huneeus, Kroft, and Lim (2021) also use Shapley values to decompose earnings inequality outcomes across multiple sources in their counterfactual analysis.

[43] For example, if the active set is $\{1,3,4,5,6\}$, employers form priors based on all initial observables and update their beliefs from post-Ph.D. publications, but workers are randomly assigned their first job.

TABLE 4. Shapley Values of Employer Learning and Initial Information

Outcome	Employer Learning	Initial Sorting	Initial Observables			
			Female	Bachelor	Ranking	Publications
Mean Publications per Worker-Year						
All	0.0314	0.0261	0.0001	0.0001	0.0001	0.0017
	14.28%	11.87%	0.06%	0.05%	0.06%	0.78%
Academia	0.1373	0.1493	0.0007	0.0007	0.0004	0.0076
	20.28%	22.04%	0.10%	0.10%	0.06%	1.12%
Industry Non-top	-0.0136	-0.0132	-5e-5	-4e-5	-0.0001	-0.0007
	-31.86%	-30.94%	-0.12%	-0.10%	-0.19%	-1.75%
Industry Top	0.0405	0.0110	0.0001	1e-5	0.0004	0.0022
	17.24%	4.69%	0.03%	0.01%	0.17%	0.93%
Total Publications per Year						
Academia	337.3	350.0	1.7	1.6	1.2	18.5
	21.04%	21.83%	0.11%	0.10%	0.07%	1.15%
Industry Non-top	-87.1	-82.8	-0.3	-0.2	-0.5	-4.8
	-32.34%	-30.74%	-0.12%	-0.09%	-0.19%	-1.80%
Industry Top	103.0	26.4	0.2	0.0	1.0	5.7
	17.17%	4.39%	0.03%	-0.01%	0.16%	0.94%
Employment at Top Firms						
Top 15% at t=10	0.0339	-0.0014	-0.0001	-0.0001	0.0003	0.0027
	12.49%	-0.50%	-0.04%	-0.02%	0.12%	1.00%
Bottom 85% at t=10	-0.0075	-0.0001	3e-5	-3e-5	-0.0002	-0.0004
	-3.29%	-0.04%	0.01%	-0.01%	-0.08%	-0.19%
Employment at Top Firms | First Job at Non-top						
Top 15% at t=10	0.0538	-0.0095	9e-7	0.0001	0.0005	0.0033
	19.96%	-3.51%	0.00%	0.02%	0.20%	1.23%
Bottom 85% at t=10	-0.0092	0.0001	0.0001	-3e-5	-0.0003	-0.0003
	-4.38%	0.07%	0.02%	-0.02%	-0.13%	-0.14%
Employment in Academia						
Top 15% at t=10	0.0551	0.0505	0.0004	0.0003	0.0002	0.0044
	18.60%	17.05%	0.12%	0.09%	0.08%	1.50%
Bottom 85% at t=10	-0.0060	-0.0100	-3e-5	-5e-5	1e-5	-0.0006
	-3.55%	-5.87%	-0.02%	-0.03%	0.01%	-0.34%

Notes: This table shows the estimated Shapley values (equation 13) of each mechanism. For each outcome, the first row reports the Shapley value in levels and the second row in percent, measured relative to the mean of the outcome when all mechanisms are active. Total publications per year are computed by rescaling the simulated totals so that the number of simulated observations matches the panel of 26,528 Ph.D. computer scientists who graduated between 2000 and 2021. Because the allocation of workers across groups varies across counterfactuals, the percentage changes in total publications differ from those in per-worker output. Top 15% and Bottom 85% refer to the distribution of true ability θ.

academia (-3.55%).

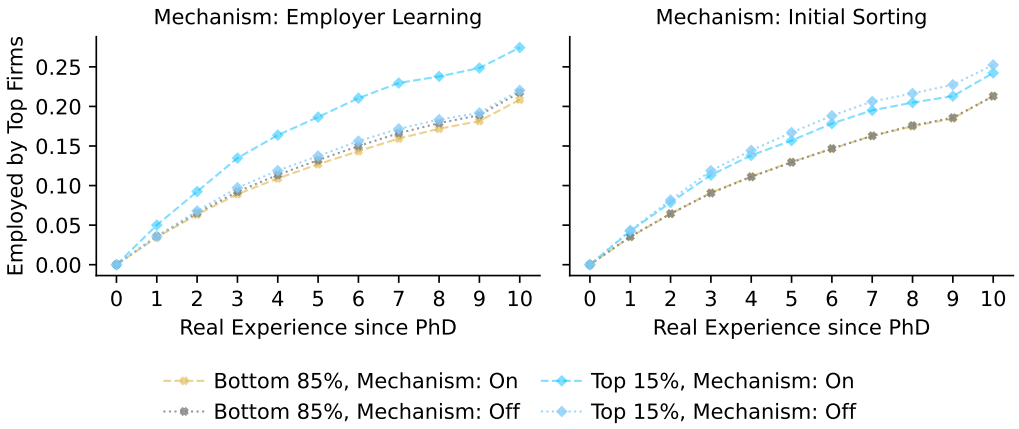
The second most important mechanism is initial sorting. If workers are instead allocated randomly across groups at entry, overall publication output per worker-year decreases by 11.87% (second column of Table 4). Initial sorting raises publication output in academia by 22%, or 350 more publications per year, and at top firms by 5%; together these more than offset the loss at non-top firms. The larger contribution in academia reflects stronger state dependence there: workers starting in academia are higher-ability on average (Table A.2), so randomizing their initial placement substantially lowers publication output.

The most important initial characteristic is the number of publications before the Ph.D. Including it in priors improves overall publication output by 0.78% and raises the employment of top 15% workers at top firms by 1% and in academia by 1.5%. Other observables, including Ph.D. ranking, matter little, with Shapley values below 0.20% in magnitude.

5.4.1 Employer Learning versus Initial Sorting

Figure 10 shows the share of workers who start at non-top firms and are employed at top firms at each t, with employer learning and initial sorting each turned on versus off.

FIGURE 10. Counterfactual Upward Mobility of Workers starting at Non-top Firms



Notes: This figure shows the share of workers employed by top firms (T) by ability group and experience, among workers starting at industry non-top firms at $t = 0$. Dashed lines average across the $2^5 = 32$ sets of mechanisms in which the panel's mechanism is active; dotted lines average across the 32 sets in which it is off. Top 15% and bottom 85% refer to the distribution of true ability θ. The mechanism is employer learning from post-Ph.D. publications in the left panel and initial sorting in the right panel.

In the left panel, shutting down learning (dotted lines) reduces the share of the top 15% at top firms by 2.5 p.p. at experience $t = 2$ and by 5 p.p. at $t \geq 5$. The share of the bottom 85% at top firms is slightly higher without post-Ph.D. learning, but as Table 4 shows, having more lower-ability workers at top firms lowers overall publication productivity.

In the right panel of Figure 10, turning off initial sorting has little impact on the upward mobility of the bottom 85%, but increases that of the top 15% by roughly 1 p.p. at $t \geq 5$. This small increase arises because random assignment places more higher-ability workers at non-top firms at entry, so the pool starting at non-top firms is more positively selected and more of them subsequently move up to top firms.

I further compare the contributions of these two mechanisms at each experience level. The Shapley value of post-Ph.D. employer learning for the publication productivity of the top 15% (left panel of Figure OA.11) rises from about 3.3% at $t = 2$ to 19.8% at $t = 10$, as employers accumulate more signals and beliefs move further from the prior. The contribution of initial sorting starts higher, at about 12%, but begins to decline after $t = 5$. The fading value of initial sorting for publication and sorting outcomes (right panel of Figure OA.11; Figure OA.12) echoes the canonical findings in Farber and Gibbons (1996) and Altonji and Pierret (2001) that wages become less correlated with the information available at labor market entry as experience accumulates. For the bottom 85%, employer learning contributes negatively to publication output, as it moves them away from top firms and academia. The contribution of initial sorting again fades with experience.

The Shapley decomposition shows that employer learning from on-the-job publications accounts for 14% of the publication output of computer scientists, a larger contribution than that of initial sorting. Even for a highly educated group whose ability might be expected to be revealed by the education system (Arcidiacono, Bayer, and Hizmo 2010), there is much for the labor market to learn after entry.[44] Taken together, these results suggest that employer learning is a substantial source of productivity gains in the first decade after the Ph.D., operating through the reallocation of higher-ability workers to employers that provide more publishing opportunities.

[44] Research success in CS requires keeping up with rapid technological advances in AI, and graduating from a top school is less predictive of research success than in economics (Footnote 3).

6 Conclusion

This paper provides empirical evidence of employer learning and quantifies its impact on sorting and allocative efficiency in the labor market for computer scientists. Combining Ph.D. dissertations, public LinkedIn profiles, and post-Ph.D. publication records, I examine how employers react to public signals of worker research ability.

Publishing at major CS conferences significantly increases the inter-firm mobility of workers at non-top firms and the probability of moving to a top tech firm. Event-study analysis further shows that the impact of a publication is immediate and persistent, and stronger for scarcer signals such as first-authored papers. These empirical patterns support the predictions of an equilibrium search model with employer learning.

I estimate an extended structural version of the model in which academia, non-top firms, and top firms provide workers with different opportunities to publish. Based on the structural estimates, the research output of early-career computer scientists would drop by 14% if the labor market could not update beliefs about worker ability from publications. Initial sorting and other initial observables matter less, suggesting that much of worker ability is revealed on the job rather than at entry. The search model with employer learning developed here can be adapted to study information frictions and (mis)allocation in other labor markets: even for a high-skilled group like Ph.D. computer scientists, uncertainty about worker ability is prevalent in the first few years of a career and, if resolved slowly, can generate inefficient allocation of labor in the longer term.

The findings also speak to a current shift in industry. In the large language model era, industry labs have become more reluctant to publish at academic conferences, reducing workers' opportunities to send public signals of research ability. Through the lens of the model, fewer opportunities to publish imply that higher-ability workers who do not start at top employers are more likely to stay hidden, lowering allocative efficiency. The implications for the value of credentials and for inequality across workers are questions for future research.

References

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

ALTONJI, JOSEPH G. AND CHARLES R. PIERRET (2001): "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116 (1), 313–350.

AMATO, NANCY M., OVIDIU DAESCU, MARIA GINI, GILLIAN R. HAYES, KATE LARSON, WEI LI, MING LIN, DAN LOPRESTI, KATIE A. SIEK, EUGENE H. SPAFFORD, AND BENJAMIN ZORN (2025): "Hiring and Promotion Guidance for Computing Researchers," Tech. rep., Computing Research Association.

ARCIDIACONO, PETER, PATRICK BAYER, AND AUREL HIZMO (2010): "Beyond Signaling and Human Capital: Education and the Revelation of Ability," *American Economic Journal: Applied Economics*, 2 (4), 76–104.

BAGGER, JESPER, FRANÇOIS FONTAINE, FABIEN POSTEL-VINAY, AND JEAN-MARC ROBIN (2014): "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, 104 (6), 1551–96.

BAGGER, JESPER AND RASMUS LENTZ (2019): "An Empirical Model of Wage Dispersion with Sorting," *The Review of Economic Studies*, 86 (1), 153–190.

BYRD, RICHARD H, PEIHUANG LU, JORGE NOCEDAL, AND CIYOU ZHU (1995): "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, 16 (5), 1190–1208.

CAHUC, PIERRE, FABIEN POSTEL-VINAY, AND JEAN-MARC ROBIN (2006): "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 74 (2), 323–364.

CARD, DAVID, ANA RUTE CARDOSO, JOERG HEINING, AND PATRICK KLINE (2018): "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36 (S1), S13–S70.

COMPUTING RESEARCH ASSOCIATION (1999): "Best Practices Memo: Evaluating Computer Scientists and Engineers for Promotion and Tenure," Tech. rep., Computing Research Association.

ERNST, MICHAEL (2013): "Evaluating Computer Science Research for Promotion and Tenure: Conference vs. Journal Publications," Academic advisory letter on the status of conference publishing in Computer Science.

FARBER, HENRY S. AND ROBERT GIBBONS (1996): "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 111 (4), 1007–1047.

GIBBONS, ROBERT AND LAWRENCE KATZ (1992): "Does Unmeasured Ability Explain Inter-Industry Wage Differentials," *The Review of Economic Studies*, 59 (3), 515–535.

GIBBONS, ROBERT AND MICHAEL WALDMAN (1999): "A Theory of Wage and Promotion Dynamics Inside Firms," *The Quarterly Journal of Economics*, 114 (4), 1321–1358.

GRÖMPING, ULRIKE (2007): "Estimators of Relative Importance in Linear Regression Based on Variance Decomposition," *The American Statistician*, 61 (2), 139–147.

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

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

JOLIVET, GRÉGORY, FABIEN POSTEL-VINAY, AND JEAN-MARC ROBIN (2006): "The empirical content of the job search model: Labor mobility and wage distributions in Europe and the US," *European Economic Review*, 50 (4), 877–907.

KAHN, LISA B. (2013): "Asymmetric Information between Employers," *American Economic Journal: Applied Economics*, 5 (4), 165–205.

KAHN, LISA B. AND FABIAN LANGE (2014): "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures," *The Review of Economic Studies*, 81 (4), 1575–1613.

KATZ, LAWRENCE F. (1985): "Three Essays on Labor Market Dynamics," Phd dissertation, Massachusetts Institute of Technology, chapter 2: Layoffs, Uncertain Recall, and the Duration of Unemployment: A Theoretical Analysis.

KLEVEN, HENRIK, CAMILLE LANDAIS, AND JAKOB EGHOLT SØGAARD (2019): "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 11 (4), 181–209.

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

LENTZ, RASMUS (2010): "Sorting by search intensity," *Journal of Economic Theory*, 145 (4), 1436–1452, search Theory and Applications.

LENTZ, RASMUS, JONAS MAIBOM, AND ESPEN R. MOEN (2026): "Directedness in Search," (26022).

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

LIM, HYUNKYEONG (2025): "Grading Policies and College Major Choice with Ability Learning," Working paper.

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

LINDENLAUB, ILSE AND FABIEN POSTEL-VINAY (2023): "Multidimensional Sorting under Random Search," *Journal of Political Economy*, 131 (12), 3497–3539.

MACROTRENDS LLC (2026): "IBM Market Cap 2012-2026 | IBM," Accessed: June 13, 2026.

MARX, MATT AND EMMA SCHARFMANN (2024): "Does Patenting Promote the Progress of Science? Evidence from Patent-Paper Pairs," Working Paper.

MILGROM, PAUL R. AND ROBERT J. WEBER (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50 (5), 1089–1122.

MISEROCCHI, FRANCESCA, SAVANNAH NORAY, AND ALICE WU (2026): "The Race between Academia and Industry for AI Researchers," Discussion Paper 26106, ROCKWOOL Foundation Berlin.

PAGE, LAWRENCE, SERGEY BRIN, RAJEEV MOTWANI, AND TERRY WINOGRAD (1999): "The PageRank Citation Ranking: Bringing Order to the Web." Tech. rep., Stanford InfoLab.

PINKSTON, JOSHUA C. (2009): "A Model of Asymmetric Employer Learning with Testable Implications," *The Review of Economic Studies*, 76 (1), 367–394.

POSTEL–VINAY, FABIEN AND JEAN–MARC ROBIN (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70 (6), 2295–2350.

RAGHAVAN, SRIRAM, MUKESH KHARE, AND JAY GAMBETTA (2025): "The 2024 IBM Research annual letter," IBM Research Blog, accessed: June 13, 2026.

SARSONS, HEATHER (2017): "Recognition for Group Work: Gender Differences in Academia," *American Economic Review*, 107 (5).

SCHÖNBERG, UTA (2007): "Testing for Asymmetric Employer Learning," *Journal of Labor Economics*, 25 (4), 651–691.

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

SORKIN, ISAAC (2018): "Ranking Firms Using Revealed Preference," *The Quarterly Journal of Economics*, 133 (3), 1331–1393.

TERVIÖ, MARKO (2009): "Superstars and Mediocrities: Market Failure in the Discovery of Talent," *The Review of Economic Studies*, 76 (2), 829–850.

VRETTAS, GEORGE AND MARK SANDERSON (2015): "Conferences versus journals in computer science," *Journal of the Association for Information Science and Technology*, 66 (12), 2674–2684.

WALDMAN, MICHAEL (1984): "Job Assignments, Signalling, and Efficiency," *The RAND Journal of Economics*, 15 (2), 255–267.

ZHU, CIYOU, RICHARD H BYRD, PEIHUANG LU, AND JORGE NOCEDAL (1997): "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Transactions on Mathematical Software (TOMS)*, 23 (4), 550–560.

Appendix Table A.1. Moments for Identifying Model Parameters

Empirical Moments	Parameters Identified
Step I. Publication Output by Entrants at $t = 0$	
Publications (Y) by Group $k \in \{A, I, T\}$	
$E[Y\|k, t=0]$, $E[Y^2\|k, t=0]$, $\Pr(Y > 0\|k, t=0)$	(N_{A0}, N_{I0}, N_{T0}); prior parameters a, b, α_k, β_k in (10)
Relationship between Publications and Observables X	
OLS coefficients on X in regressions of $\text{asinh}(Y)$ on X and dummies for k	Coefficients a, b in prior (10)
OLS coefficients on X in regressions of $1[Y > 0]$ on X and dummies for k	Coefficients a, b in prior (10)
Step II. Publication Output and Labor Market Transitions at $t = 0, 1, \ldots, 10$	
Publications	
OLS coefficients from regressions of Y on a polynomial of experience (cubic for T, quadratic for A, linear for I)	Changes in publication opportunities N_{kt} with experience (11)
$E[Y\|k]$ pooled over $t \in \{0, \ldots, 10\}$, for $k \in \{I, T\}$	Parameters in Step I and in (11)
Origin-specific Job-to-Job Transitions	
Downward J2J from each rung in I	$F_I(j)$
Upward J2J from each rung in I	$F_I(j)$, $\Lambda_{II,0}$
J2J to top firms (T) from each rung in I	$\Lambda_{IT,0}$
Downward J2J from each rung in A	$F_A(j)$
Upward J2J from each rung in A	$F_A(j)$, $\Lambda_{AA,0}$ for non-postdocs, $\Lambda^p_{AA,0}$ for postdocs
J2J to non-top industry firms (I) from each rung in A	$\Lambda_{AI,0}$, $\Lambda^p_{AI,0}$
J2J to top firms (T) from each rung in A	$\Lambda_{AT,0}$, $\Lambda^p_{AT,0}$
Coefficients from Regressions of Mobility on Publication Output	
From Industry Non-top (I)	
J2J to top firms on $\text{asinh}(Y)$ and interaction with t, for workers in I	$\Lambda_{IT,1}$
J2J to top firms on $1[Y > 0]$, for workers in I, pooled over t	$\Lambda_{IT,1}$
Upward J2J within I on $\text{asinh}(Y)$ and interaction with t	$\Lambda_{II,1}$
From Academia	
Upward J2J within A on $\text{asinh}(Y)$ and interaction with postdoc status	$\Lambda_{AA,1}$, $\Lambda^p_{AA,1}$
Between Academia and Industry	
J2J from industry (I/T) to academia on $\text{asinh}(Y)$	$\Lambda_{IA,1}$
J2J to top firms on $\text{asinh}(Y)$ and interaction with postdoc status, for workers in A	$\Lambda_{AT,1}$, $\Lambda^p_{AT,1}$
J2J to top firms on $1[Y > 0]$, among postdocs	$\Lambda^p_{AT,1}$

Appendix Table A.2. Estimated Parameters (Step I)

parameter	Estimated via SMM	Estimate	SE
Publication Opportunities			
(N_{A0}, N_{I0}, N_{T0})	X	(4,2,3)	
Shape Parameters in (10)			
a_0 (α_I non-top)	X	-2.3596***	0.1948
a_{female}	X	-0.3820**	0.1789
$a_{bachelor}$	X	-0.2192	0.3166
a_{rank}	X	-0.3217*	0.1648
$a_{pre-pubs}$	X	3.2008***	0.7665
a_A ($\alpha_A - \alpha_I$, acad rel. to non-top)	X	0.6229***	0.1662
a_T ($\alpha_T - \alpha_I$, top rel. to non-top)	X	0.3263	0.2015
b_0 (β_I Non-top)	X	-0.0726	0.2176
b_{female}	X	-0.1550	0.3647
$b_{bachelor}$	X	-0.2230	0.5197
b_{rank}	X	-0.1760	0.2782
$b_{pre-pubs}$	X	0.2185	0.4523
b_A ($\beta_A - \beta I$, acad rel. to non-top)	X	-0.0133	0.2256
b_T ($\beta_T - \beta_I$, top rel. to non-top)	X	-0.0391	0.2615
Effect on Log-Odds of Prior Mean			
$j(i,0) \in A$		0.6362***	0.1238
$j(i,0) \in I$		-2.2869***	0.0951
$j(i,0) \in T$		0.3654***	0.1091
female		-0.2269	0.2428
bachelor in the US		0.0038	0.2320
Ph.D. ranking		-0.1457	0.1334
Papers before PhD		2.9824***	0.3337

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Appendix Table A.3. Estimated Parameters (Step II)

Parameter	SMM	Estimate	SE
Changes of Publication Opportunities with Experience (11)			
ν_{A1} on t	X	0.2332	0.7313
ν_{A2} on t^2	X	-0.0182	0.1004
ν_{I1} on t	X	-0.0693	See Figure 7
$(\nu_{T1}, \nu_{T2}, \nu_{T3})$ on (t, t^2, t^3)	X	(0.3191, -0.0812, 0.0048)	
Exogenous Separations			
Layoff δ		0.0035	
Involuntary J2J $(\kappa_I, \kappa_T, \kappa_A)$		(0.0500, 0, 0.0228)	
Exits $(\mu, \overline{\mu})$		(0.0176, 0.0601)	
Job Arrival Rates: Origin = Industry			
$\Lambda_{II,0}$	X	0.1913***	0.0062
$\Lambda_{II,1}$	X	-0.9527***	0.2913
$\Lambda_{IT,0}$	X	0.0281***	0.0018
$\Lambda_{IT,1}$	X	1.3593***	0.4044
$\Lambda_{IA,0} = \Lambda_{TA,0}$		0.0134	
$\Lambda_{IA,1} = \Lambda_{TA,1}$	X	2.0467***	0.7917
$(\Lambda_{TI,0}, \Lambda_{TI,1})$		(0.0450, 0)	
Job Arrival Rates: Origin = Academia			
$\Lambda_{AA,0}$		0.0797	
$\Lambda_{AA,1}$	X	-0.1738	0.4557
$\Lambda_{AI,0}$		0.0090	
$\Lambda_{AI,1}$	X	-0.2404	0.8221
$\Lambda_{AT,0}$		0.0023	
$\Lambda_{AT,1}$	X	0.4865	2.1855
Postdoc premiums			
$\Lambda^p_{AA,0}$	X	0.0521**	0.0235
$\Lambda^p_{AA,1}$	X	2.4417*	1.4473
$\Lambda^p_{AI,0}$	X	0.1671***	0.0097
$\Lambda^p_{AI,1}$	X	-0.0731	0.5752
$\Lambda^p_{AT,0}$	X	0.0210***	0.0073
$\Lambda^p_{AT,1}$	X	2.7337	3.8978
Job Arrival Rates: Origin = Unemployment			
Λ_{0A}		0.0943	
Λ_{0I}		0.1079	
Λ_{0T}		0.0182	
Offer distribution: $F_k(j) = \sum_{j' \leq j} f_k(j')$, $f_k(j) = \frac{exp(v_k(j))}{\sum_{j'} exp(v_k(j'))}$			
$v_A(0) = v_I(0)$		0	
$v_A(1)$	X	-1.4590	4.7849
$v_A(2)$	X	-1.2235	2.7252
$v_A(3)$	X	0.3052	0.2444
$v_I(1)$	X	-1.8747***	0.4883
$v_I(2)$	X	-0.5711***	0.1060
$v_I(3)$	X	-1.3385***	0.4913
$v_I(4)$	X	-1.7624**	0.7928
$v_I(5)$	X	0.1025	0.0811

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Online Appendix

Appendix A. Additional Results

Table OA.1. Descriptive Statistics: Matched Computer Scientists

	All		Estimation Sample	
	Mean	SD	Mean	SD
Gender from Name or Picture				
Female 0.130	0.336	0.127	0.334	
Male	0.707	0.455	0.708	0.455
Education				
Year of Ph.D.	2011.435	5.868	2011.466	5.746
Ph.D. in CS (∋ EECS)	0.531	0.499	0.530	0.499
Ph.D. in EE	0.469	0.499	0.470	0.499
Bachelor in the U.S.	0.399	0.490	0.377	0.485
Bachelor in Asia	0.509	0.500	0.532	0.499
Research Outputs Post Ph.D.				
Papers Post PhD	2.357	8.662	2.574	9.080
First-Authored Papers	0.336	1.318	0.364	1.381
Paper-Patent Matches 0.251	1.577	0.281	1.656	
Any Paper	0.290	0.454	0.310	0.462
Any First-Authored Paper	0.150	0.357	0.161	0.367
Any Paper-Patent Match	0.076	0.265	0.084	0.278
Employment Post Ph.D.				
Yrs with Full-time Employment	10.857	5.850	11.162	5.713
Tenure-track Employers	0.258	0.560	0.257	0.563
Postdoc Employers	0.181	0.426	0.198	0.443
Top Firms	0.330	0.562	0.373	0.584
Nontop Firms	1.529	1.374	1.548	1.373
Observations	30,871		26,528	

Notes: This table summarizes the sample of Ph.D. computer scientists whose doctoral dissertations can be matched to a LinkedIn profile with employment records after the Ph.D. The first two columns summarize the full sample of graduates from the top 60 CS departments in the U.S. between 2000 and 2021. I use the full sample throughout Section 4. The estimation sample comprises those who have a job record in the year after the Ph.D. and are ever employed by a firm in the largest connected component of employers; this restriction is imposed because the PageRank value of a firm, which is used to group employers in the structural estimation, is identified only within the largest connected component. I use the estimation sample in the structural analysis in Section 5.

Table OA.2. Descriptive Statistics: Worker-Year Panel

Currently employed by:	Nontop Firms		Top Firms		Academia	
	Mean	SD	Mean	SD	Mean	SD
Experience (yrs since Ph.D.)	7.051	5.180	6.731	4.942	6.682	5.233
Current Position						
Tenure-track	0.000	0.011	0.000	0.015	0.667	0.471
Postdoc	0.000	0.000	0.008	0.090	0.163	0.370
Research Scientist	0.228	0.419	0.291	0.454	0.082	0.275
Engineer	0.514	0.500	0.636	0.481	0.036	0.186
Manager	0.148	0.355	0.178	0.383	0.013	0.114
Senior Position	0.489	0.500	0.367	0.482	0.045	0.207
Research Outputs						
Any paper	0.029	0.168	0.114	0.318	0.201	0.401
Asinh(paper)	0.034	0.213	0.146	0.447	0.290	0.644
Any first-authored paper	0.009	0.093	0.034	0.181	0.047	0.211
Any paper with a matched patent	0.009	0.093	0.037	0.190	0.015	0.121
Transitions between t and $t+1$						
New employer next yr	0.130	0.337	0.069	0.253	0.101	0.302
Employed by top firms next yr	0.022	0.148	0.948	0.222	0.008	0.092
Research scientist at top firms next yr	0.005	0.067	0.010	0.097	0.004	0.060
Engineer at top firms next yr	0.015	0.121	0.023	0.151	0.004	0.063
Move to a higher-wage firm	0.035	0.183	0.033	0.179	0.031	0.172
Move to a higher-wage position	0.045	0.207	0.023	0.149	0.057	0.232
Promotion to senior roles	0.038	0.192	0.042	0.201	0.007	0.086
Observations	194,209		54,252		86,719	

Notes: This table summarizes the worker×year panel for the matched Ph.D.s. The first two columns report means across worker×year observations for workers currently employed by a non-top firm in industry. The second pair covers workers at top firms, and the third workers in academia (including postdocs, tenure-track positions, and other roles).

Table OA.3. Examples of CS Papers and Matched Patent Applications

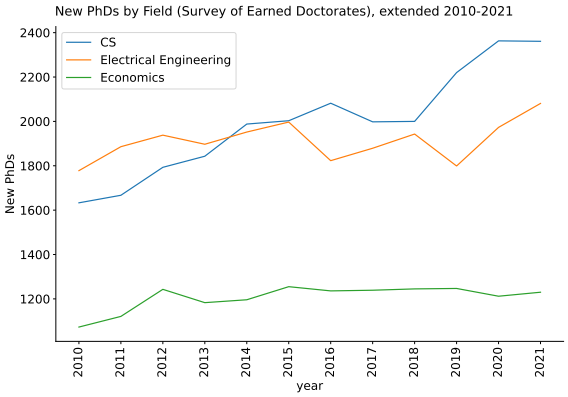
			Papers		Matched Patent Applications		
Firm	Team	Text Dist.	Title	M/Yr	Title	Filing M/Yr	Published M/Yr
Yahoo	100%	0.233	UNBIASED ONLINE ACTIVE LEARNING IN DATA STREAMS	08/2011	ONLINE ACTIVE LEARNING IN USER-GENERATED CONTENT STREAMS	10/2011	05/2013
Adobe	80%	0.273	FORECASTING HUMAN DYNAMICS FROM STATIC IMAGES	07/2017	FORECASTING MULTIPLE POSES BASED ON A GRAPHICAL IMAGE	04/2017	10/2018
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" 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 Dist." 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 V3 model ("text-embedding-3-small"). 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.

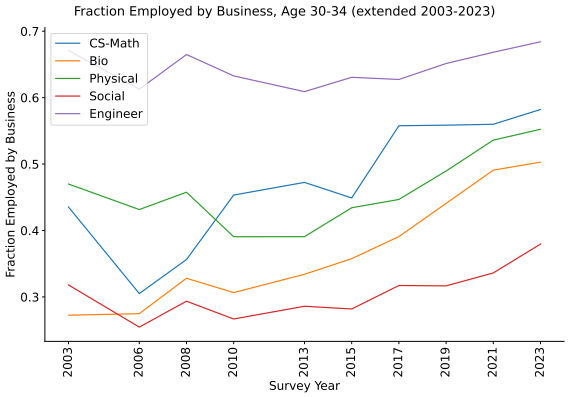
Figures

Figure OA.1. CS PhDs in NSF Surveys

(a) Num. New PhDs



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



Notes: Panel (a) plots the number of new Ph.D.s by discipline in the Survey of Earned Doctorates. Panel (b) shows the share of graduates employed by business between age 30-34, using data from the Survey of Doctoral Recipients.

Figure OA.2. Research Scientist Job Postings

(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

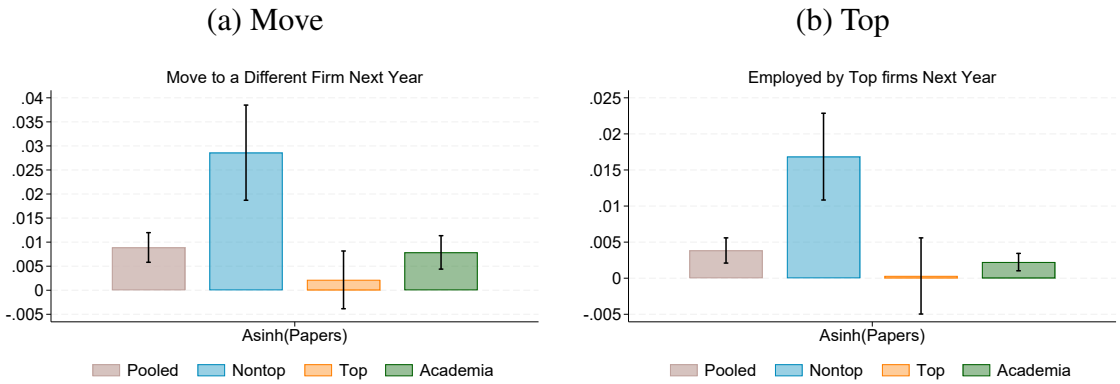
Minimum qualifications:

- 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:

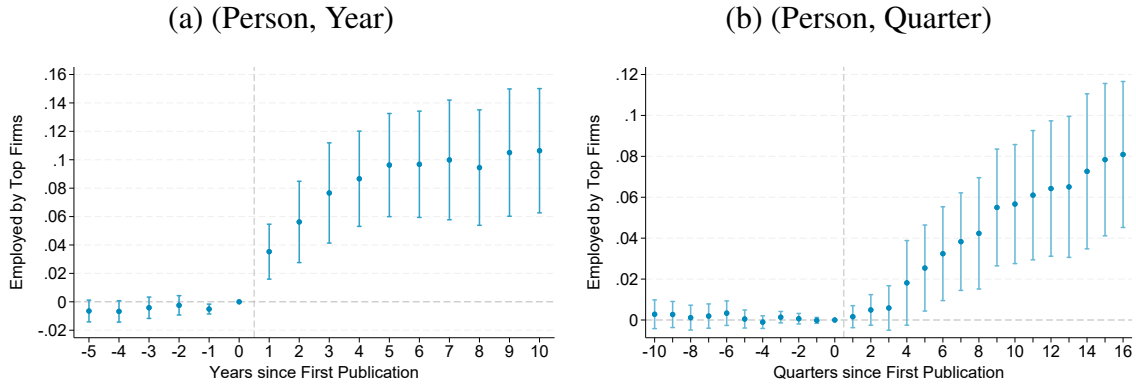
- Relevant work experience, including full time industry experience or as a re
- Strong publication record
- Ability to design and execute on research agenda.

Figure OA.3. Publication output – asinh(papers) on job outcomes next year



Notes: This figure repeats Figure 1, measuring publication output by the inverse hyperbolic sine of the number of publications this year rather than by an indicator for any publication. The outcomes are (a) moving to a different employer and (b) moving to a top firm in the next year, and the specification is the one described under Figure 1. Each panel reports the pooled estimate and the estimates separately by whether workers are at non-top firms, top firms, or in academia this year. 95% confidence intervals are computed from robust standard errors clustered at the Ph.D. school×cohort level.

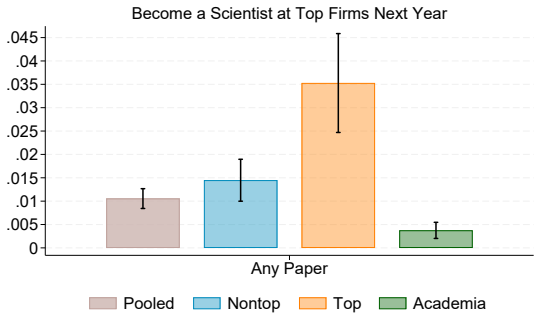
Figure OA.4. Employment by Top Firms among Workers who Publish at Non-top Firms at 0



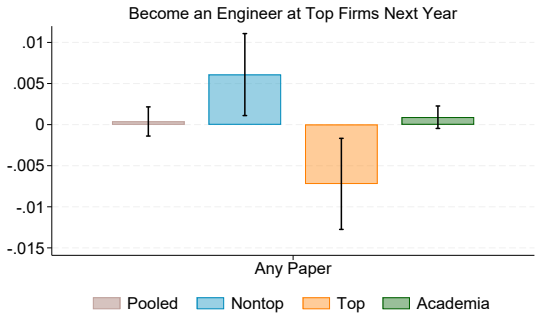
Notes: This figure shows the difference-in-differences estimates for the dynamic effects of the first publication 1-3 years after Ph.D. on employment by top firms, without controlling for worker fixed effects in equation (8). Instead, the regressions control for origin firm fixed effects ($j(i,0)$ at event time), a quartic polynomial in experience, position characteristics, demographics and Ph.D. school × cohort fixed effects. The estimation sample is the same as for Figure 2.

Figure OA.5. Additional Employment Outcomes in $t+1$ on Any Paper in t

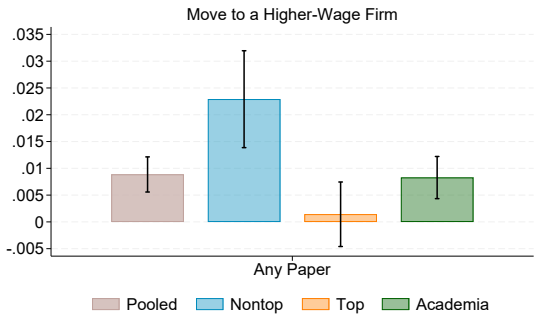
(a) Scientist at Top



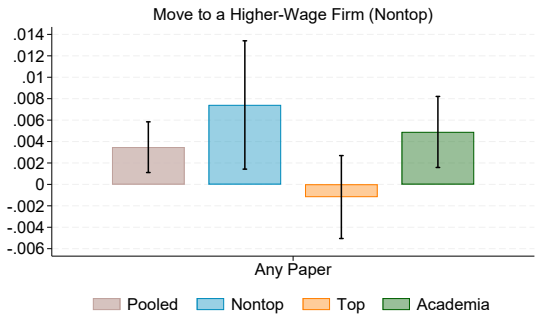
(b) Engineer at Top



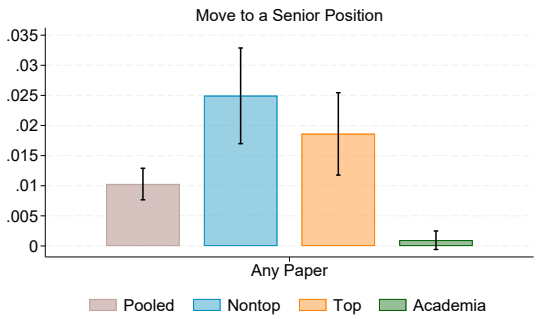
(c) Higher-wage Firm



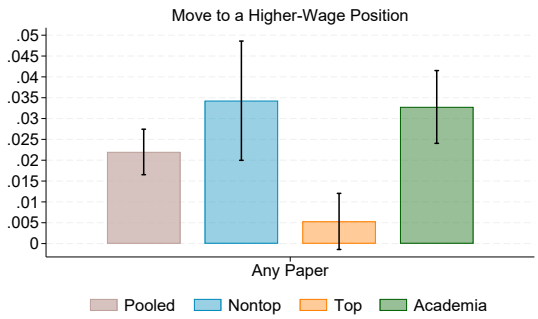
(d) Higher-wage Firm (Nontop)



(e) Senior Position

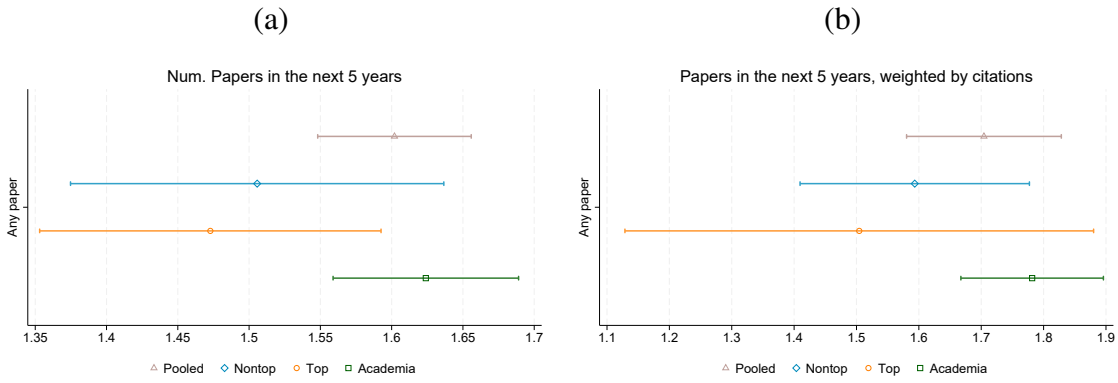


(f) Higher-wage Position



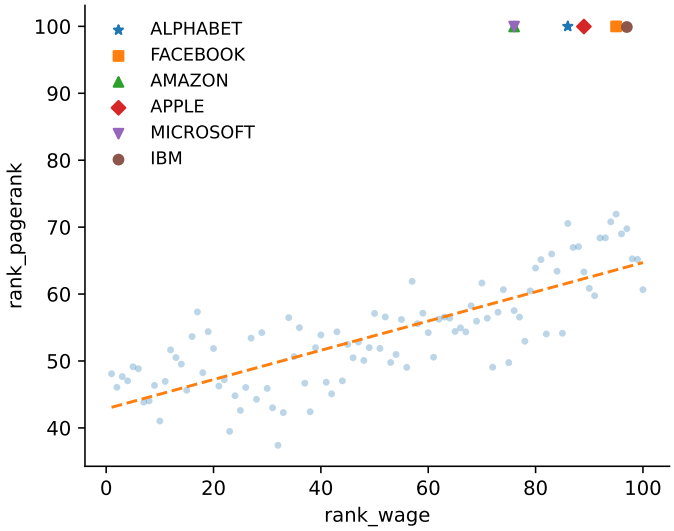
Notes: This figure shows the predictive effect of any publication this year on additional job outcomes in the next year, using the regression described under Figure 1. Panels (a) and (b) define the outcomes as *becoming* a research scientist or an engineer at a top firm; among workers currently at top firms, panel (a) is estimated on those not yet in a research role and (b) on those not yet in an engineering role. Panels (c) measures moves to a higher-wage firm, using wages posted for foreign workers in H-1B and green card petitions. Panel (d) excludes transitions to top firms from moves to a higher-wage firms. Panels (e)-(f) measure promotion to a more senior job title or to a higher-wage position. Each panel reports the pooled estimate and the estimates separately by whether the worker is currently at a non-top firm, a top firm, or in academia.

Figure OA.6. Future Publications in $[t+1, t+5]$ on Any Papers in t



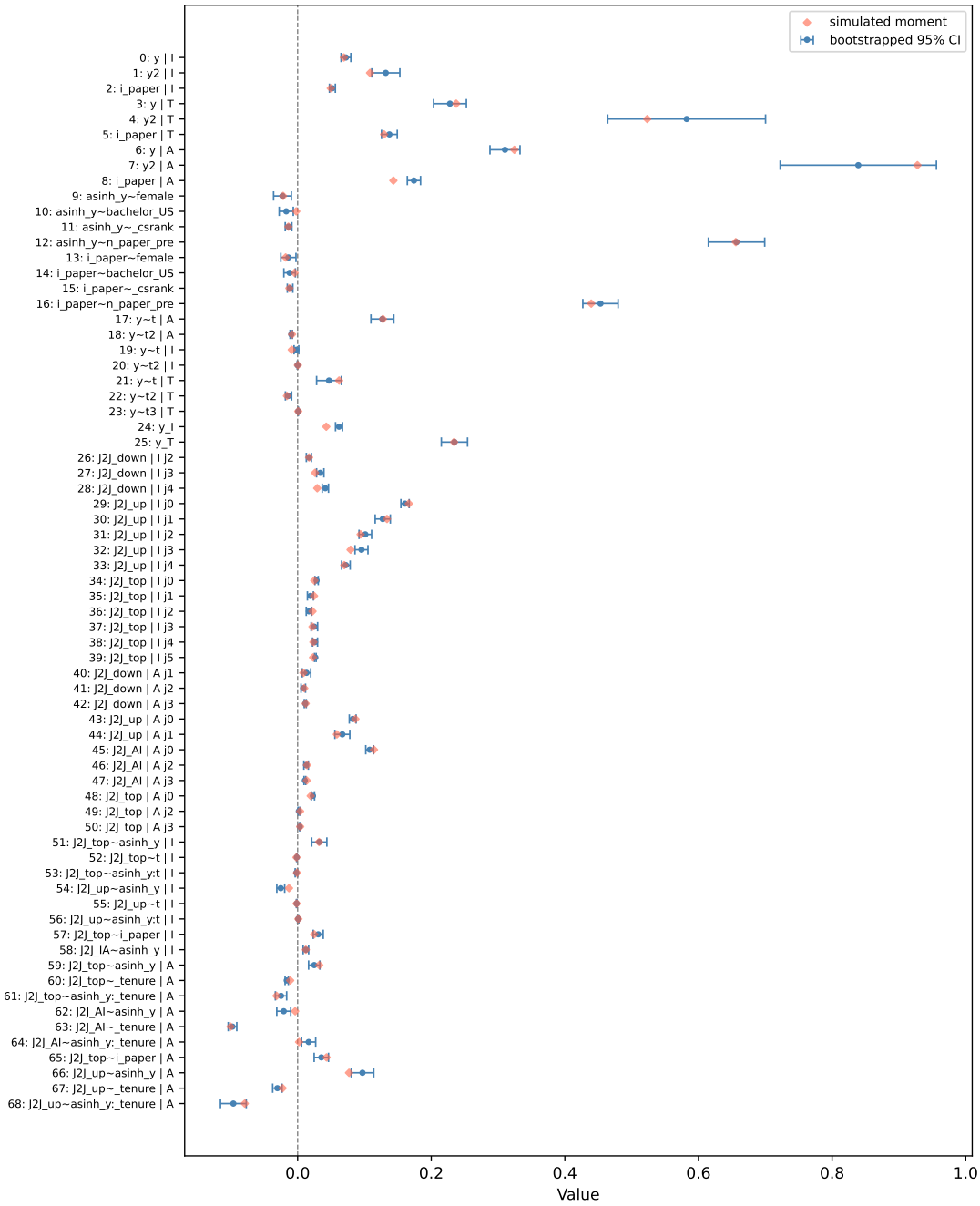
Notes: This figure shows the predictive effect of any publication this year on publications in the next five years. The regression model is explained in detail under Figure 1. Panel (a) shows the estimates for the total number of papers in the next five years, and panel (b) shows the estimates for the citation-weighted publication counts. For each publication I use the citation that it will receive within one year.

Figure OA.7. Ranking Firms: PageRank versus Wages



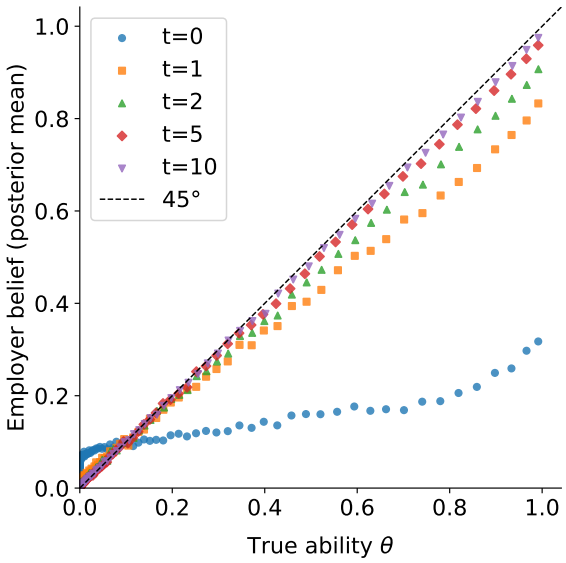
Notes: This figure shows a binned scatter plot of the percentile rank of firms by PageRank value against their percentile rank by mean wages, computed from posted wages for foreign workers. The six top firms (T) are shown individually. Their PageRank values place them at the top of the distribution, above what their wage ranks alone would imply.

Figure OA.8. Empirical versus Simulated Moments



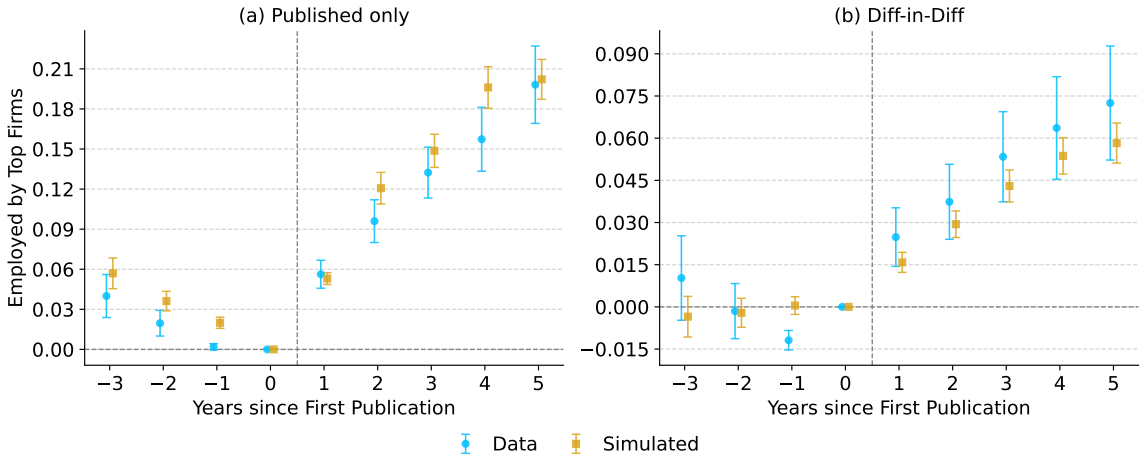
Notes: The 95% confidence interval around each empirical moment is constructed from 1,000 bootstrapped sample in the training set.

Figure OA.9. Employer Beliefs versus True Ability



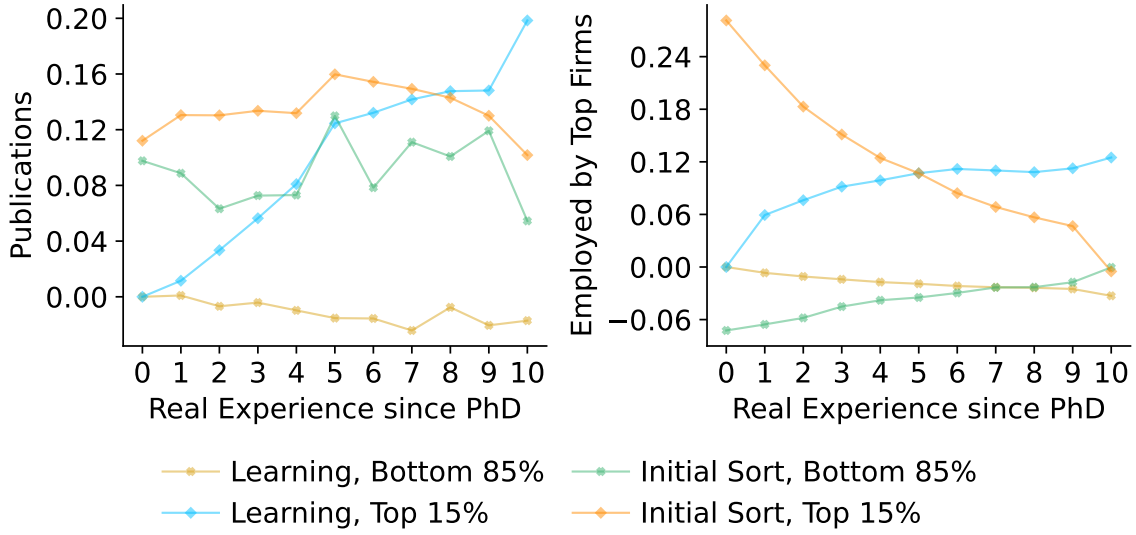
Notes: This figure is a binned scatter plot of the posterior mean of employer beliefs against true ability θ in the simulated panel, at different levels of real experience t. Given belief Beta(α, β), the posterior mean is $\frac{\alpha}{\alpha+\beta}$.

Figure OA.10. Employment by Top Firms around First Publication



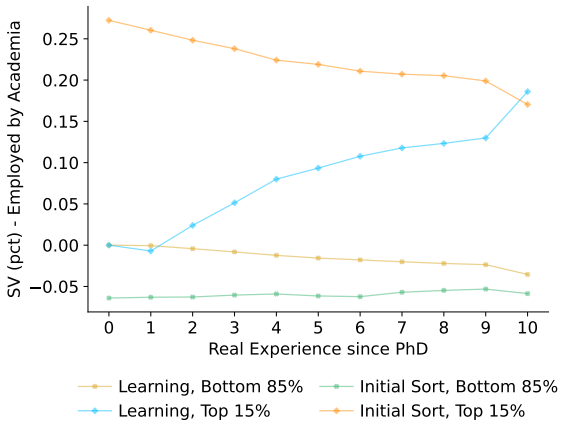
Notes: Panel (a) shows the event-study estimates for employment by top firms around the first publication after the Ph.D., among workers employed at non-top firms when publishing their first paper (event year 0). Panel (b) shows the dynamic difference-in-differences estimates. The control group comprises workers without a publication after the Ph.D.; I use an ordered logit to assign each a placebo event year, conditional on initial characteristics.

Figure OA.11. Shapley Values (Relative to Benchmark) by Experience



Notes: This figure shows the estimated Shapley values of employer learning and initial sorting by real experience, separately for workers in the top 15% and bottom 85% of the distribution of true ability. The outcome in the left panel is the mean number of publications per worker-year; in the right panel, the share of workers employed by top firms. Shapley values are expressed as a fraction of the benchmark mean outcome when all mechanisms are active.

Figure OA.12. Shapley Values (Relative to Benchmark) for Employment in Academia



Notes: This figure shows the estimated Shapley values of employer learning and initial sorting for the share of workers employed in academia, by real experience, separately for workers in the top 15% and bottom 85% of the distribution of true ability θ. Shapley values are expressed as a fraction of the mean outcome when all mechanisms are active.

Appendix B. Benchmark Model

The benchmark model is introduced in Section 2. This appendix derives the flow-balance equations in steady state and proves the equilibrium results: Propositions 1 and Predictions 1, 2 and 3.

B.1 Value Function and Flow-balance Equations

Joint value is increasing in firm productivity.

Lemma 1 establishes that the joint value defined by (3) is strictly increasing in firm productivity. A worker therefore accepts an outside offer if and only if the poaching firm is more productive: voluntary turnover is efficient and moves workers up the job ladder. This justifies the integral limits in (3), the poaching probability in the flow-balance equations below, and the revealed-preference interpretation of job-to-job flows in Section 5.

Lemma 1 *Assume the mutually exclusive shocks satisfy* $\mu+\delta+\kappa+\lambda_1(E[\theta|S,t])<1$ *at every state* (S,t), $\mu+\lambda_0<1$, *and the offer distribution* F *has a continuous density* f. *Then the joint value* $V(\phi,S,t)$ *is differentiable and strictly increasing in* ϕ *at every* (S,t), *with*

$$\frac{\partial V(\phi,S,t)}{\partial \phi}=\frac{\partial E[m(\phi,\theta)|S,t]}{\partial \phi} + \sum_{y=0}^{N} Pr(y|S,t)\cdot\frac{1-\widetilde{\delta}-\gamma\lambda_1\left(\frac{\alpha_0+S+y}{\alpha_0+\beta_0+(t+1)N}\right)\overline{F}(\phi)}{1+\rho}\,\frac{\partial V(\phi,S+y,t+1)}{\partial \phi} \tag{A.1}$$

where $\widetilde{\delta}:=\mu+\delta+\kappa$. *The derivative is bounded:* $\frac{\partial V(\phi,S,t)}{\partial \phi}\in\left[1+\eta N\frac{\alpha_0+S}{\alpha_0+\beta_0+tN},\ \frac{(1+\rho)(1+\eta N)}{\rho+\widetilde{\delta}}\right]$.

Proof:

The expected flow payoff $E[m(\phi,\theta)|S,t]=\phi\,(1+\eta NE[\theta|S,t])$ is bounded by $\overline{\phi}(1+\eta N)$ and strictly increasing in ϕ, with $\frac{\partial}{\partial\phi}E[m(\phi,\theta)|S,t]=1+\eta N\frac{\alpha_0+S}{\alpha_0+\beta_0+tN}$.

Existence and uniqueness. A *candidate* $(U^{(n)},V^{(n)})$ assigns a bounded value to unemployment in any state (S,t) and to employment in any state (ϕ,S,t). Denote by $\mathscr{T}$ the mapping from a candidate to the right-hand sides of (2) and (3) at every state; the equilibrium values are, by definition, its fixed point.$\mathscr{T}$ satisfies Blackwell's sufficient conditions on bounded candidates with the sup norm over all states: (1) it is monotone (collecting terms on the RHS of (2) and (3), every coefficient on the candidate is nonnegative), and (2) adding a constant c to the candidate adds $\frac{1-\mu}{1+\rho}\cdot c$. $\mathscr{T}$ is therefore a contraction with modulus

$\frac{1-\mu}{1+\rho} < 1$. Starting from $(U^{(0)}, V^{(0)}) \equiv (0,0)$ and letting $(U^{(n+1)}, V^{(n+1)}) = \mathscr{T}(U^{(n)}, V^{(n)})$, the iterates converge uniformly to the unique bounded fixed point (U,V) satisfying (2) and (3).

Differentiability. Suppose $V^{(n)}(\cdot,S,t)$ is continuously differentiable for every (S,t) (true at $n=0$). For a differentiable candidate $v := V^{(n)}(\cdot, S+y, t+1)$, the poaching term in (3) differentiates as $\frac{d}{d\phi}\int_{\phi}^{\overline{\phi}} \left(v(\widetilde{\phi}) - v(\phi)\right) dF(\widetilde{\phi}) = -\overline{F}(\phi)\, v'(\phi)$. Hence $V^{(n+1)}(\cdot,S,t)$ is continuously differentiable, with

$$\frac{\partial V^{(n+1)}(\phi,S,t)}{\partial \phi} = (1+\eta NE[\theta|S,t]) + \sum_{y=0}^{N} Pr(y|S,t)\, \frac{1-\widetilde{\delta} - \gamma\lambda_1(\frac{\alpha_0+S+y}{\alpha_0+\beta_0+(t+1)N})\overline{F}(\phi)}{1+\rho} \cdot \frac{\partial V^{(n)}(\phi, S+y, t+1)}{\partial \phi}. \tag{A.2}$$

Equation (A.2) is an affine recursion in the derivatives with coefficients that are nonnegative and sum to at most $\frac{1-\widetilde{\delta}}{1+\rho} < 1$. The sequence $\left\{\frac{\partial}{\partial\phi}V^{(n)}\right\}$ is therefore itself generated by a contraction with modulus $\frac{1-\widetilde{\delta}}{1+\rho}$ and converges uniformly over all (ϕ,S,t) to the unique bounded solution of (A.1). Since $V^{(n)} \to V$ uniformly and $\frac{\partial}{\partial\phi}V^{(n)}$ converges uniformly, the limit $V(\cdot,S,t)$ is continuously differentiable with derivative equal to the limit (Theorem 7.17 in Rudin 1976), establishing (A.1).

Monotonicity and bounds. Starting from $\frac{\partial}{\partial\phi}V^{(0)} \equiv 0$, every iterate of (A.2) is nonnegative by induction, since the recursion has nonnegative intercept and nonnegative coefficients; the uniform limit therefore satisfies $\frac{\partial V(\phi,S,t)}{\partial\phi} \geq 0$ at every state. Substituting $\frac{\partial V(\phi,S+y,t+1)}{\partial\phi} \geq 0$ into (A.1) yields $\frac{\partial V(\phi,S,t)}{\partial\phi} \geq 1+\eta NE[\theta|S,t] > 0$: the joint value is strictly increasing in ϕ at every state. Taking the supremum over states in (A.1) and using coefficients at most $\frac{1-\widetilde{\delta}}{1+\rho}$ gives $\sup\frac{\partial}{\partial\phi}V \leq (1+\eta N) + \frac{1-\widetilde{\delta}}{1+\rho}\sup\frac{\partial}{\partial\phi}V$, hence $\sup\frac{\partial}{\partial\phi}V \leq \frac{(1+\rho)(1+\eta N)}{\rho+\widetilde{\delta}}$. □

Flow-balance Equations

In steady state, the inflow to unemployment equals the outflow. Denote by $u(\theta,S,t)$ the density of workers with ability θ and public state (S,t). Entrants are characterized by $(S,t) = (0,0)$ and because they are assumed to start with a job, $u(\theta,0,0) = 0$. At any $t \geq 1$,

the flow-balance equation is:

$$\underbrace{(\mu+\lambda_0)\cdot u(\theta,S,t)}_{\text{outflow}} = \underbrace{\delta \sum_{y=0}^{\min(N,S)} Pr(Y=y|\theta)\cdot l(\theta,S-y,t-1)}_{\text{inflow from empl}}. \tag{A.3}$$

The outflow comprises unemployed workers who either exit the market for good (with probability μ) or receive an outside offer (with probability λ_0). The inflow comprises workers who in the previous period are employed with experience $t-1$, produce y signals such that the market belief is updated to Beta$(\alpha_0+S,\beta_0+N\cdot t-S)$, and are laid off after the belief update.

Denote by $l(\theta,S,t)$ the density of workers with ability θ and public state (S,t). At $t=0$, given population distribution $\theta\sim$ Beta(α_0,β_0), $l(\theta,0,0)\propto\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}$. At $t\geq 1$, the flow-balance equation for the employed with public state (S,t) and true ability θ is:

$$\begin{aligned} l(\theta,S,t) &= \lambda_0 u(\theta,S,t) + (1-\mu-\delta)\sum_{y=0}^{\min(N,S)} Pr(Y=y|\theta)\cdot l(\theta,S-y,t-1) \\ &= \left(1-\mu-\frac{\mu\delta}{\mu+\lambda_0}\right)\sum_{y=0}^{\min(N,S)} Pr(Y=y|\theta)\cdot l(\theta,S-y,t-1) \end{aligned} \tag{A.4}$$

All currently employed workers in state (S,t) constitute the outflow, as the state advances to $(S+y,t+1)$. The inflow on the right-hand side comprises the unemployed whose state remains unchanged at (S,t), and the employed previously in state $(S-y,t-1)$, which is updated to (S,t) upon producing y publications. Workers who experience the advance layoff notice or κ shock remain employed and hence stay within $l(\theta,S,t)$. The second equation is established by plugging in (A.3).

Lemma 2 $l(\theta,S,t)=\left(1-\mu-\frac{\mu\delta}{\mu+\lambda_0}\right)^t\cdot\binom{tN}{S}\theta^S(1-\theta)^{tN-S}\cdot l(\theta,0,0)$.

Proof:

(Base) Consider $l(\theta,S,t)$ at $t=1$:

$$\begin{aligned} \forall S\in[0,N]:\ l(\theta,S,1) &= \left(1-\mu-\frac{\mu\delta}{\mu+\lambda_0}\right) Pr(Y=S|\theta)\cdot l(\theta,0,0) \\ &= \left(1-\mu-\frac{\mu\delta}{\mu+\lambda_0}\right)\binom{N}{S}\theta^S(1-\theta)^{N-S}\cdot l(\theta,0,0). \end{aligned}$$

(Inductive step) Suppose at some $t \geq 1$, $l(\theta,S,t) = \left(1-\mu-\frac{\mu\delta}{\mu+\lambda_0}\right)^t \cdot \binom{tN}{S} \cdot \theta^S(1-\theta)^{t\cdot N-S} \cdot l(\theta,0,0)$. Then by the flow-balance equation (A.4),

$$\begin{aligned} l(\theta,S,t+1) &= \left(1-\mu-\frac{\mu\delta}{\mu+\lambda_0}\right) \cdot \sum_{y=0}^{\min(N,S)} Pr(Y=y|\theta) \cdot l(\theta,S-y,t) \\ &= \left(1-\mu-\frac{\mu\delta}{\mu+\lambda_0}\right)^{t+1} \cdot \underbrace{\sum_{y=0}^{\min(N,S)} \binom{N}{y} \cdot \binom{tN}{S-y}}_{(*)} \cdot \theta^S(1-\theta)^{(t+1)N-S} \cdot l(\theta,0,0) \end{aligned}$$

By Vandermonde's identity, $(*) = \binom{(t+1)N}{S}$. Therefore,

$$l(\theta,S,t+1) = \left(1-\mu-\frac{\mu\delta}{\mu+\lambda_0}\right)^{t+1} \cdot \binom{(t+1)N}{S} \theta^S(1-\theta)^{(t+1)N-S} \cdot l(\theta,0,0).$$

□

Define $G(\phi,\theta,S,t)$ as the density of workers who have ability θ, are characterized by state (S,t) and are employed working for firms with productivity $\leq \phi$. In steady state, the flow-balance condition is:

$$\begin{aligned} G(\phi,\theta,S,t) = &\underbrace{\lambda_0 F(\phi) \cdot u(\theta,S,t)}_{\text{from unempl}} + \underbrace{\kappa \cdot F(\phi) \sum_{y=0}^{\min(N,S)} Pr(y|\theta)\, l(\theta,S-y,t-1)}_{\text{from involuntary J2J}} \\ &+ \underbrace{(1-\widetilde{\delta}-\lambda_1(E[\theta|S,t]) \cdot \overline{F}(\phi)) \sum_{y=0}^{\min(N,S)} Pr(y|\theta)\, G(\phi,\theta,S-y,t-1)}_{\text{retained from the employed}} \end{aligned} \tag{A.5}$$

where the exogenous separation risk $\widetilde{\delta} := \mu+\delta+\kappa$. Dividing the equation above by $l(\theta,S,t)$ yields the conditional distribution of firms:

$$\begin{aligned} G(\phi|\theta,S,t) = &F(\phi)\frac{\lambda_0\frac{\delta}{\mu+\lambda_0}+\kappa}{1-\mu-\frac{\mu\delta}{\mu+\lambda_0}} \\ &+ (1-\widetilde{\delta}-\lambda_1(E[\theta|S,t]) \cdot \overline{F}(\phi)) \sum_{y=0}^{\min(N,S)} \frac{Pr(y|\theta) \cdot l(\theta,S-y,t-1)}{l(\theta,S,t)} G(\phi|\theta,S-y,t-1) \end{aligned} \tag{A.6}$$

At $t=0$, entrants draw offers from the sampling distribution F, so $G(\phi|0,0)=F(\phi)$. At $t>1$, $G(\phi|S,t)<F(\phi)$.

Lemma 3 *Conditional on public state (S,t), the distribution of firms is independent of worker ability θ: $G(\phi|\theta,S,t)\equiv G(\phi|S,t)$.*

Proof:
(Base) At $t=0$, $G(\phi|\theta,0,0)\equiv F(\phi)$. At $t=1$,

$$G(\phi|\theta,S,1)=F(\phi)\frac{\frac{\lambda_0}{\mu+\lambda_0}\delta+\kappa}{1-\mu-\frac{\mu}{\mu+\lambda_0}\delta}+\left(1-\widetilde{\delta}-\lambda_1(E[\theta|S,1])\overline{F}(\phi)\right)\frac{Pr(S|\theta)\,l(\theta,0,0)}{l(\theta,S,1)}F(\phi).$$

By Lemma 2, $\frac{Pr(S|\theta)\times l(\theta,0,0)}{l(\theta,S,1)}=(1-\mu-\frac{\mu\delta}{\mu+\lambda_0})^{-1}$ is independent of ability θ. So $G(\phi|\theta,S,1)=G(\phi|S,1)$.

(Inductive step) Suppose at some $t\geq 1$, $G(\phi|\theta,S,t)=G(\phi|S,t)$. Then for any θ,θ', using the flow-balance equation (A.6),

$$\begin{aligned}
&G(\phi|\theta,S,t+1)-G(\phi|\theta',S,t+1)\\
&\quad=\left(1-\widetilde{\delta}-\lambda_1(E[\theta|S,t+1])\overline{F}(\phi)\right)\\
&\qquad\times\sum_{y=0}^{\min(N,S)}G(\phi|S-y,t)\left(\frac{Pr(y|\theta)l(\theta,S-y,t)}{l(\theta,S,t+1)}-\frac{Pr(y|\theta')l(\theta',S-y,t)}{l(\theta',S,t+1)}\right)
\end{aligned}$$

in which the ratio $\frac{Pr(y|\theta)l(\theta,S-y,t)}{l(\theta,S,t+1)}=(1-\mu-\frac{\mu\delta}{\mu+\lambda_0})^{-1}\cdot\frac{\binom{N}{y}\binom{tN}{S-y}}{\binom{(t+1)N}{S}}$ is independent of ability θ. Therefore, $G(\phi|\theta,S,t+1)-G(\phi|\theta',S,t+1)=0$. □

Lemma 3 establishes equation (5) for $G(\phi|S,t)$ in the main text.

B.2 Proof of Proposition 1

The key assumption for Proposition 1 is that on-the-job offer arrival rate λ_1 is increasing in posterior mean $E[\theta|S,t]$ when employers observe S publications by workers with t years of real experience.

Proof of Proposition 1(a): $S > S' \Rightarrow G(\phi|S,t) < G(\phi|S',t)$.

(Base) At $t = 1$, by Lemma 2, $\frac{Pr(S|\theta)\, l(\theta,0,0)}{l(\theta,S,1)} = \left(1-\mu-\frac{\mu\delta}{\mu+\lambda_0}\right)^{-1}$, independent of signals S. Thus the difference satisfies

$$\frac{G(\phi|S,1)}{F(\phi)} - \frac{G(\phi|S',1)}{F(\phi)} = -\left(\lambda_1(E[\theta|S,1]) - \lambda_1(E[\theta|S',1])\right) \cdot \frac{\overline{F}(\phi)}{\left(1-\mu-\frac{\mu\delta}{\mu+\lambda_0}\right)}. \quad \text{(A.7)}$$

$S > S'$ implies the posterior mean $E[\theta|S,1] = \frac{\alpha_0+S}{\alpha_0+\beta_0+N} > E[\theta|S',1]$. Under the assumption that $\lambda_1' > 0$, (A.7) is negative.

(Inductive step) At $t \geq 2$, suppose $G(\phi|S,t-1) < G(\phi|S',t-1)$ for any $S > S'$. By equation (A.6), the difference between $G(\phi|S,t)$ and $G(\phi|S',t)$ satisfies:

$$G(\phi|S,t) - G(\phi|S',t) = \left(1-\widetilde{\delta}-\lambda_1(E[\theta|S,t])\cdot\overline{F}(\phi)\right) \sum_{y=0}^{\min(N,S)} \frac{\binom{N}{y}\binom{(t-1)N}{S-y}}{\binom{tN}{S}} G(\phi|S-y,t-1)$$
$$- \left(1-\widetilde{\delta}-\lambda_1(E[\theta|S',t])\cdot\overline{F}(\phi)\right) \sum_{y=0}^{\min(N,S')} \frac{\binom{N}{y}\binom{(t-1)N}{S'-y}}{\binom{tN}{S'}} G(\phi|S'-y,t-1)$$

Under the assumption that λ_1 is increasing in the posterior mean, $\left(1-\widetilde{\delta}-\lambda_1(E[\theta|S,t])\cdot\overline{F}(\phi)\right) < \left(1-\widetilde{\delta}-\lambda_1(E[\theta|S',t])\cdot\overline{F}(\phi)\right)$ for any $S > S'$ and $\phi < \overline{\phi}$. It therefore suffices to show that $\sum_{y=0}^{\min(N,S)} \frac{\binom{N}{y}\binom{(t-1)N}{S-y}}{\binom{tN}{S}} G(\phi|S-y,t-1)$ is decreasing in S.

The weight $\frac{\binom{(t-1)N}{S-y}\binom{N}{y}}{\binom{tN}{S}}$ is the probability of drawing $(S-y)$ "good" items among S draws without replacement from tN items, of which $(t-1)N$ are "good" and N are "bad". Let X_S denote the number of good items in S such draws. I construct a coupling of X_S and $X_{S'}$ for $S' < S$. Place the tN items in a random sequence, and let z_n indicate whether the n-th item is good. Define

$$\widetilde{X}_{S'} = \sum_{k \leq S'} z_k, \qquad \widetilde{X}_S = \widetilde{X}_{S'} + \sum_{k=S'+1}^{S} z_k.$$

These have the same marginal distributions as $X_{S'}$ and X_S, and $Pr(\widetilde{X}_S \geq \widetilde{X}_{S'}) = 1$, so $X_S \succ^{FOSD} X_{S'}$ by Strassen's theorem (Strassen 1965).

By the inductive hypothesis $G(\phi|S,t-1) < G(\phi|S',t-1)$ for $S' < S \leq (t-1)N$, so

$\overline{G}(\phi|x,t-1) := 1-G(\phi|x,t-1)$ is increasing in x. Using the expectation characterization of $X_S \succ^{FOSD} X_{S'}$,

$$\sum_{x=0}^{S} Pr(X_S = x)\,\overline{G}(\phi|x,t-1) > \sum_{x=0}^{S} Pr(X_{S'} = x)\,\overline{G}(\phi|x,t-1),$$

which, writing $x = S-y$, is equivalent to

$$\sum_{y=0}^{\min(N,S)} \frac{\binom{N}{y}\binom{(t-1)N}{S-y}}{\binom{tN}{S}} G(\phi|S-y,t-1) < \sum_{y=0}^{\min(N,S')} \frac{\binom{N}{y}\binom{(t-1)N}{S'-y}}{\binom{tN}{S'}} G(\phi|S'-y,t-1).$$

In combination with λ_1 increasing in the posterior mean, this implies $G(\phi|S,t) < G(\phi|S',t)$, completing the induction. □

Proof of Proposition 1(b): $\theta > \theta' \Rightarrow G(\phi|\theta,t) < G(\phi|\theta',t)$.

By Lemma 3,

$$G(\phi|\theta,t) = \sum_{S=0}^{tN} Pr(S|\theta,t)\cdot G(\phi|\theta,S,t) = \sum_{S=0}^{tN} Pr(S|\theta,t)\cdot G(\phi|S,t) \tag{A.8}$$

Proposition 1(a) has established that $G(\phi|S,t)$ is decreasing in signals S. It suffices to show that $S|(\theta,t) \succ^{FOSD} S|(\theta',t)$ for any $\theta > \theta'$ at any $t \geq 1$.

Denote by S_θ a random variable with distribution Binomial(tN,θ) given any $t \geq 1$. I construct a coupling for S_θ and $S_{\theta'}$ where $\theta > \theta'$. Consider tN independent $Uniform(0,1)$ variables, denoted by $u_1,...,u_{tN}$. Define $\widetilde{S}_\theta = \sum_{i=1}^{tN} 1[u_i \leq \theta]$. Since $\theta > \theta'$, $u_i \leq \theta'$ implies $u_i \leq \theta$. So $\widetilde{S}_{\theta'} \leq \widetilde{S}_\theta$ almost surely. By Strassen's theorem, $S_\theta \succ^{FOSD} S_{\theta'}$.

So for any $\theta > \theta'$, $S_\theta \succ^{FOSD} S_{\theta'}$, and since $\overline{G} = 1-G$ is increasing in S,

$$\begin{aligned}\theta > \theta' \Rightarrow & \sum_{S=0}^{tN} Pr(S|\theta,t)\cdot \overline{G}(\phi|S,t) > \sum_{S=0}^{tN} Pr(S|\theta',t)\cdot \overline{G}(\phi|S,t) \\ \Rightarrow & G(\phi|\theta,t) < G(\phi|\theta',t).\end{aligned}$$

□

B.3 Theoretical Motivations for $\lambda_1' > 0$

Model 1 – Endogeneous Search Efforts

Following Bagger and Lentz (2019), suppose job offers for the currently employed are generated through a choice of search intensity, e, at cost $c(e) = \frac{\zeta}{2}e^2$. The arrival rate of outside offers is $\lambda_1(e) = \bar{\lambda}_1 \cdot e$ with $\bar{\lambda}_1 > 0$. TThe search intensity maximizes the joint value:

$$V(\phi,S,t) = \max_{e\geq 0} \sum_{y=0}^{N} Pr(y|S,t) \cdot \frac{\lambda_1(e)}{1+\rho} \underbrace{\gamma \int_{\phi}^{\overline{\phi}} (V(\phi',S+y,t+1) - V(\phi,S+y,t+1)) dF(\phi')}_{\text{expected surplus from bargaining with a higher-}\phi\text{ firm}} - c(e)$$
$$+ \text{other terms in (3), independent of } e \tag{A.9}$$

Taking the FOC w.r.t. search effort e,

$$e^*(\phi,S,t) = \frac{\bar{\lambda}_1}{\zeta} \frac{\gamma}{1+\rho} \sum_{y=0}^{N} Pr(y|S,t) \int_{\phi}^{\overline{\phi}} (V(\phi',S+y,t+1) - V(\phi,S+y,t+1)) dF(\phi').$$

Lentz (2010) shows that when the production technology is supermodular, search intensity is increasing in worker skill. In this model, the flow value (equation 1) is supermodular, and so is the expected flow $E[m(\phi,\theta)|S,t] = \phi(1+\eta \cdot N\frac{\alpha+S}{\alpha+\beta+Nt})$ in ϕ and S. Workers with more publication records (hence more favorable market belief) search more in expectation of higher match value at more productive firms. The assumption that λ_1 is increasing with posterior mean is therefore justified in a model with endogenous search intensity. A model with endogenous search intensity can therefore rationalize the assumption that λ_1 increases with the posterior mean.

Model 2 – Two-dimensional Firm Productivity

Suppose firms vary by baseline productivity ϕ and the opportunities to publish N. Following Gibbons and Waldman (1999) and Gibbons and Katz (1992), I assume the flow output is additive in baseline production and research production:

$$m(\phi,N,\theta) = \phi + \eta(N) \cdot \theta \tag{A.10}$$

where $\eta(N)$ represents returns to publications and $\eta'(N) > 0$. Consider two types of firms, 1 and 2, with $\phi_1 > \phi_2$, $N_1 < N_2$. Assuming efficient turnover conditional on public information, workers with prior Beta(α, β) at firm 1 accept offers from firm 2 if the posterior mean satisfies:

$$\frac{\alpha + y}{\alpha + \beta + N_1} > \frac{\phi_1 - \phi_2}{\eta(N_2) - \eta(N_1)}. \tag{A.11}$$

Based on the inequality above, holding the belief precision $\alpha + \beta$ fixed, there are thresholds at $k = 0, ..., N_1$ defined as:

$$r_k(\alpha + \beta) \coloneqq (1 + \frac{N_1}{\alpha + \beta}) \cdot \frac{\phi_1 - \phi_2}{\eta(N_2) - \eta(N_1)} - \frac{k}{\alpha + \beta}. \tag{A.12}$$

The thresholds decrease in k and summarize which workers move from firm 1 to firm 2 when offers arrive:

$$\frac{\alpha}{\alpha + \beta} > r_k(\alpha + \beta) \;\Rightarrow\; \forall y \geq k : \frac{\alpha + y}{\alpha + \beta + N_1} > \frac{\phi_1 - \phi_2}{\eta(N_2) - \eta(N_1)}.$$

Suppose offers arrive with a constant probability λ_1. The probability that a worker with ability θ and employer belief Beta(α, β) moves from firm 1 to firm 2 voluntarily is:

$$\lambda_1 \cdot \sum_{y=0}^{N_1} Pr(y|N_1, \theta) \times 1\left[\frac{\alpha}{\alpha + \beta} > r_y(\alpha + \beta)\right],$$

which is non-decreasing in the prior mean $\frac{\alpha}{\alpha+\beta}$, holding the belief precision $\alpha + \beta$ and ability θ fixed. Workers with a more favorable employer belief are more likely to move to firms that provide more opportunities to publish. Although the offer arrival rate λ_1 is held fixed, in equilibrium there are more transitions to firms with $N_2 > N_1$ from workers with higher beliefs.

B.4 Proofs of Predictions on Job Mobility

I derive three predictions on job mobility under the assumption that λ_1 is differentiable and increasing in the posterior mean of the employer belief, $E[\theta|S,t]$; Predictions 2–3 additionally assume λ_1 is locally linear over the range of posterior means induced by one period of signals. The predictions characterize job mobility in the steady-state equilibrium.

Proof of Prediction 1 (Job Mobility in Response to Publications)

Consider a pair of coworkers at some firm with productivity $\phi < \overline{\phi}$, who share the same employer belief Beta(α, β). Suppose they produce y and y' publications, with $y > y'$. The posterior mean given y is higher: $E[\theta|\alpha,\beta,y] = \frac{\alpha+y}{\alpha+\beta+N} > E[\theta|\alpha,\beta,y']$. In the next period, a worker moves to a more productive firm either upon receiving an outside offer above ϕ, with probability $\lambda_1(E[\theta|\alpha,\beta,y]) \cdot \overline{F}(\phi)$, or upon the involuntary κ shock with a draw above ϕ, with probability $\kappa \cdot \overline{F}(\phi)$. Since μ, δ, and κ do not vary with the employer belief, these channels are identical across the two workers and cancel in the comparison. The difference in job mobility, and in upward mobility, between the two workers is therefore

$$\overline{F}(\phi) \cdot \left(\lambda_1(E[\theta|\alpha,\beta,y]) - \lambda_1(E[\theta|\alpha,\beta,y'])\right),$$

which is positive since λ_1 is increasing and $\overline{F}(\phi) > 0$ for $\phi < \overline{\phi}$. Workers with more publications thus exhibit both higher job mobility and higher upward mobility than coworkers at the same firm with the same prior. □

Proof of Prediction 2 (Belief Precision)

Consider two coworkers who share the same current belief and the same N, and who produce y and y' publications, respectively, with $y > y'$. The gap in their posterior means, $\frac{y-y'}{\alpha+\beta+N}$, is smaller when the belief precision $\alpha+\beta$ is higher. The difference in J2J or upward mobility between the two workers, as derived above, is $\overline{F}(\phi) \cdot \left(\lambda_1(\frac{\alpha+y}{\alpha+\beta+N}) - \lambda_1(\frac{\alpha+y'}{\alpha+\beta+N})\right)$. By the Mean Value Theorem, there exists $\xi \in \left[\frac{\alpha+y'}{\alpha+\beta+N}, \frac{\alpha+y}{\alpha+\beta+N}\right]$ such that

$$\lambda_1\left(\frac{\alpha+y}{\alpha+\beta+N}\right) - \lambda_1\left(\frac{\alpha+y'}{\alpha+\beta+N}\right) = \lambda_1'(\xi) \times \underbrace{\frac{y-y'}{\alpha+\beta+N}}_{(*)} \tag{A.13}$$

in which the term (*) is decreasing in the belief precision $\alpha+\beta$. When λ_1 is locally linear over the relevant range, $\lambda_1'(\xi)$ is approximately constant across the two comparisons, so the difference in mobility responses is governed by the term (*). Comparing beliefs with the same mean but different precision, the mobility response to the same gap in publications is therefore smaller when the belief is more precise. □

Proof of Prediction 3 (Signal Scarcity)

By (A.13), the mobility response to an additional publication equals $\lambda_1'(\xi)$ times the change in the posterior mean, $\frac{1}{\alpha+\beta+N}$, which is decreasing in the number of opportunities to publish, N. When λ_1 is locally linear, $\lambda_1'(\xi)$ varies little across the comparison, so a lower N generates a larger change in the posterior mean and a larger mobility response. □

Appendix C. Data

C.1 Data Sources

Doctoral Dissertations. Appendix Table OC.1 displays the number of dissertations by year. For school×year cells with particularly low or missing data on ProQuest, I collected about 15,000 additional Ph.D. graduates 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, so I supplemented these with additional records from KiltHub, CMU's open-access repository. Appendix Table OC.2 provides a detailed breakdown of dissertations collected from ProQuest versus school-specific sources. The total number of Ph.D. graduates in the sample stays around 3,000–3,300 per year from the top 60 schools since 2006 (Appendix Figure OC.2).

LinkedIn Profiles With the Recruiter Lite subscription, LinkedIn allowed me to view public profiles within my third-degree connections. To expand coverage, I sent connection requests before querying to individuals likely to be linked to CS Ph.D. graduates, including authors at CS conferences, professors in CS departments, and research scientists at various companies. If an individual is on LinkedIn but falls outside my third-degree connections, the search result indicates "Out of Network." There were about 1,800 out-of-network profiles in total, out of 50,000 queries that returned at least one profile. I manually checked a random sample of these profiles and found that most had fewer than 100 connections on LinkedIn, suggesting they are not active users and are less likely to report a complete job history.

Publications Data. The main data source of research papers is Scopus, an abstract and citation database of peer-reviewed literature launched by Elsevier in 2004. For each conference/journal × year, I submitted a query via the Scopus Search API, which returns a list of papers with information such as authors, title, abstract, ISSN, DOI, number of

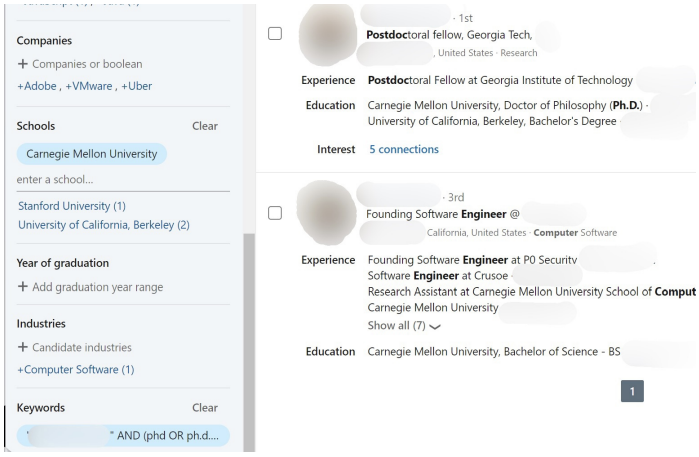
Table OC.1. Number of Profiles by Year

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

Table OC.2. Number of Profiles by School (ProQuest vs. School-specific Dissertation Database or Websites)

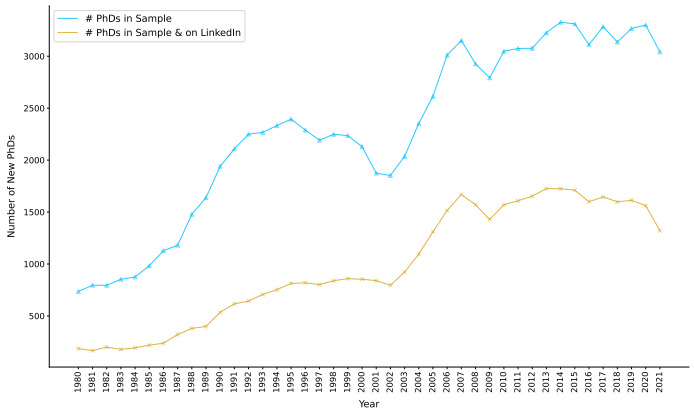
	ProQuest Dissertations			School-specific Sources		
School	# Dissertations	LinkedIn Profiles	Matched	# Dissertations	LinkedIn Profiles	Matched
austin	2028	990	845	1671	762	635
berkeley	3169	1949	1618	836	369	272
caltech	721	435	296	402	184	112
cmu	2357	1537	1259	2332	920	695
cornell	1738	962	685	481	203	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

Figure OC.1. 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 OC.2. 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 OC.2) 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.

citations, volume, issue, and publication date. Scopus also provides affiliation IDs at the paper $\times$ author level. I then submitted a query for each affiliation ID via the Affiliation Search API, which returns the corresponding institution's name and location. I consider a paper by author i affiliated with j as her on-the-job research if (1) j can be matched with an employer of i on her LinkedIn profile; and (2) author i is employed by j at the time of publication.

If a paper has multiple authors, I flag it when the majority of coauthors come from i's Ph.D. institution, which is likely to indicate a publication of her dissertation, especially if it happens within the first year after Ph.D. I also flag papers where coauthors 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 list Microsoft as her affiliation at the time of publication, but if her coauthors come from Yahoo, the work was likely done at Yahoo rather than Microsoft. Such papers typically include a note such as "This work was done when X was at ...".

To evaluate paper quality, I collected citations from Scopus, which covers both journal articles and conference papers. Citations from other conference papers are particularly important in computer science. For each paper, I counted the number of citations by year since publication, excluding self-citations.

Patents Data. To obtain a more complete picture of innovation activity, I collected data on patent applications submitted to the United States Patent and Trademark Office (USPTO) by matched computer scientists. I required the assignee (firm) on the patent application to match the inventor's employer as shown on LinkedIn, and the year of the initial filing to fall within her tenure at the firm. The main datasets I used for patent records are:

1. Patent Examination Research Dataset (PatEX) 2022, sourced from the Public Patent Application Information Retrieval (PAIR) system (Miller 2020);
2. Patent Assignment Dataset 2023 (Marco et al. 2015);
3. PatentsView December 2025 Release.

Public PAIR was retired by the USPTO in 2022. I used the 2022 release of the patent examination dataset to capture publicly-viewable patent applications that were published by the USPTO by June 2023. I merged it with Patent Assignment Dataset to identify the assignees that would own the patents if granted. Patent applications are typically filed by individuals and then assigned to their employers, who become the assignees.

Since patent applications are published with an 18-month delay by default, PAIR 2022 disproportionately misses applications filed in 2021–2022 that had not yet been made public by the USPTO at the time of the data release. To get a more complete picture of patenting activity throughout 2022, I used the PatentsView 2025 release to identify additional patent applications filed by computer scientists in my analysis sample. PatentsView also provides the most up-to-date patent abstracts, which are useful for the paper-patent matching described below.

A key advantage of using PAIR rather than PatentsView for earlier years is that PAIR contains provisional applications, which offer a low-cost way to secure an early priority date before submitting a formal patent application. Without records of provisional applications, the recorded filing date can be misleading. Inventors have a 12-month window after filing a provisional to submit a corresponding nonprovisional application claiming its priority. The nonprovisional is then published 18 months after the priority date by default. PAIR, and the PatEX dataset derived from it, was therefore particularly useful for identifying whether and when a provisional application was filed.

C.2 Matching CS Publications with Patent Applications

I matched CS publications with patent applications in four steps. First, I identified potential matches by merging the paper-author panel with the patent application-inventor panel by individual full name, allowing a $\pm$ three-year window between paper publication and patent filing. I restricted both panels to papers and patents contributed by matched computer scientists in the ProQuest-LinkedIn sample. I preserved the full author and inventor teams so that I could measure the overlap between authors on the paper and inventors on the patent application, a key predictor of a true match.

Second, I converted the title and abstract of each paper and each patent application into embeddings using OpenAI's "text-embedding-3-small", transforming text into 1536-dimensional numeric vectors. I then measured the ℓ_2 distance for each potential paper-patent pair.

Third, I built a labeled dataset to train and evaluate predictive models of paper-patent matching. I randomly drew 470 potential matches and recruited four research assistants to label whether each patent application was closely related to the findings of the corresponding CS publication. I asked the RAs to review the patent records on Google Patents or the original USPTO documents to compare the patent claims with the contributions of

the paper. In addition, I used the patent-paper pairs (PPP) dataset from Marx and Scharfmann (2024), who match publications from OpenAlex with granted patents from USPTO records. I found 42,625 CS paper-patent pairs in PPP, of which 2,400 were labeled as true matches.[45]

I estimated several models that predict a true paper-patent match using (1) the distance between paper and patent abstract embeddings, (2) the share of authors on a paper who also appear as inventors on the corresponding patent application, (3) the share of inventors who appear as paper authors, (4) team size on the paper and the patent, and the interactions between (1) and (2)-(4). The performance of each model on a held-out test set is summarized below. The most important predictors in the random forest model are the distance between embeddings and the team overlap measures (2)–(3), which are also the most significant in the regressions.

Table OC.3. Performance of Models that Predict a Paper-Patent Match

Model	auc	ypred_cutoff	TP	FP	F1
Linear	0.994416	0.300000	0.913242	0.001766	0.941176
Linear Ridge	0.994418	0.300000	0.913242	0.001766	0.941176
Linear Lasso	0.994399	0.300000	0.914764	0.001373	0.944969
Logit	0.994028	0.425000	0.933029	0.001766	0.951863
Logistic Ridge	0.993977	0.425000	0.933029	0.001766	0.951863
Logistic Lasso	0.993847	0.425000	0.934551	0.002060	0.950464
Random Forest	0.992736	0.500000	0.943683	0.000883	0.964230
Ensemble	0.994382	0.350000	0.939117	0.001471	0.957331

I used the average predicted probability across linear regressions, logistic regressions, and the random forest model as the ensemble probability score. The optimal threshold appears to be 0.350.

Applying the ensemble to the panel of ProQuest-LinkedIn computer scientists, I established a paper-patent linkage if the following conditions are satisfied:

1. The worker is listed as both an author on the paper and an inventor on the corresponding patent application.

[45] I identified CS matches based on conference name or ID associated with each publication record on OpenAlex. OpenAlex is less consistent than Scopus in its coverage of CS conference proceedings. In addition, the PPP dataset is restricted to granted patents rather than patent applications. For these reasons, I built a more comprehensive dataset of CS paper-patent application pairs using Scopus and USPTO application records, replicating the predictive models in Marx and Scharfmann (2024) as closely as possible.

2. The predicted probability from the ensemble exceeds 0.350.

Figure OC.3 shows that papers with a matched patent are of higher quality on average. In the first year, they receive roughly the same number of citations as those without a matched patent. However, the gap begins to widen around two years after the paper appears, which coincides with the disclosure of patent applications (Figure OC.4). The quality difference between papers with or without a matched patent suggests two mechanisms: first, incumbent employers have private information about a paper's quality and act on it by filing for a patent; second, it takes time for outsiders to recognize valuable research, and the divergence in citations emerges around the same time as the revelation of a matched patent.

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

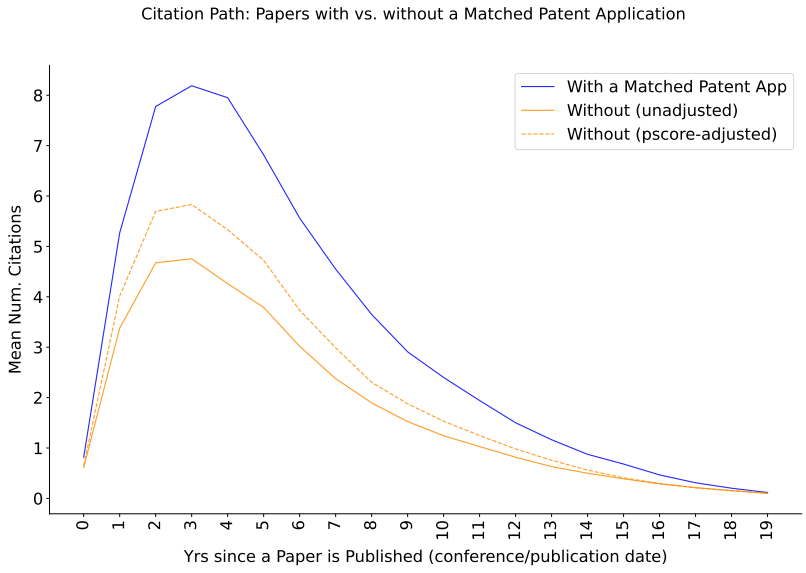
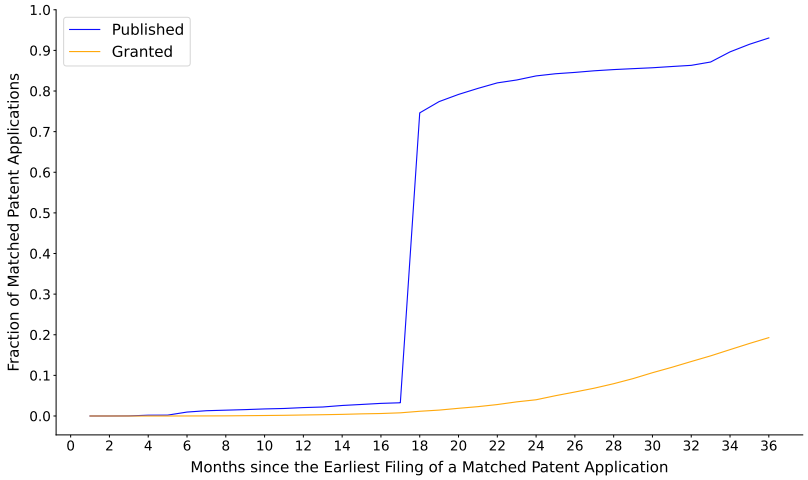


Figure OC.4. Publication of Patent Applications that are Matched to CS Papers



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 of the American Inventors Protection Act (AIPA 1999). About 20% of matched patent applications are disclosed later than 18 months. 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). 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.

Appendix D. Model Extension

This appendix presents the full details of the extended model introduced in Section 5.1. I describe the value functions, the flow-balance equations for all three employer groups and across all experience levels (including the entry condition at $t = 0$ and the terminal state at $t = 10$), and the joint distribution over firm productivity within each group.

D.1 Value functions

Since the opportunities to publish vary by employer group, cumulative publications S and experience t are no longer sufficient statistics for the state of the workers. Instead, the public state is (α, β, t), where α, β are shape parameters of the employer belief.

Unemployed

Unemployed workers with experience t and employer belief Beta(α,β) receive job offers from each group $k \in \{A,I,T\}$ with probability Λ_{0k}. The value of unemployment is:

$$\begin{aligned} U(\alpha,\beta,t) =& b + \frac{\Lambda_{0I}\,\gamma}{1+\rho}\int_{\phi\in I}\left(V_I(\phi,\alpha,\beta,t) - U(\alpha,\beta,t)\right)dF_I(\phi) \\ &+ \frac{\Lambda_{0T}\,\gamma}{1+\rho}\left(V_T(\overline{\phi},\alpha,\beta,t) - U(\alpha,\beta,t)\right) \\ &+ \frac{\Lambda_{0A}\,\gamma}{1+\rho}\int_{\phi\in A}\left(V_A(\phi,\alpha,\beta,t) - U(\alpha,\beta,t)\right)dF_A(\phi) \\ &+ \frac{1-\mu_t}{1+\rho}U(\alpha,\beta,t) \end{aligned} \tag{A.14}$$

where top firms share productivity $\phi \equiv \overline{\phi}$. Rearranging yields:

$$\begin{aligned} &(\rho+\mu_t)\cdot U(\alpha,\beta,t) \\ &= (1+\rho)b + \gamma \sum_{k\in\{A,I,T\}} \Lambda_{0k}\left(E[V_k(\phi,\alpha,\beta,t)|\phi\in k] - U(\alpha,\beta,t)\right) \end{aligned} \tag{A.15}$$

Relative to (2) in the benchmark model, the value of unemployment now accounts for transitions into each group $k \in \{A,I,T\}$ with group-specific offer arrival rates Λ_{0k}. Equation (A.15) applies at all experience levels $t \in \{1,2,\ldots,10\}$, where $\mu_t = \mu$ for $t < 10$ and $\mu_{10} = \overline{\mu}$ captures higher attrition at the terminal state.

Employed

The value function for the employed workers is group-specific. For workers with employer belief Beta(α,β) at experience t, the value of matching with a firm with productivity ϕ in

group k is:

$$
\begin{aligned}
& V_k(\phi,\alpha,\beta,t) \qquad\qquad (A.16)\\
=& \underbrace{E[m_k(\phi,\theta,t)|\theta\sim \text{Beta}(\alpha,\beta)]}_{\text{flow}}\\
&+\sum_{y=0}^{N_{kt}} Pr(y|\alpha,\beta,N_{kt})\cdot\frac{\delta+\kappa_k+\sum_{k'\neq k}\Lambda_{kk'}(\pi(y),t+1)}{1+\rho}\cdot\underbrace{U(\alpha+y,\beta+N_{kt}-y,t+1)}_{\text{unemployment}}\\
&+\sum_{y=0}^{N_{kt}}\frac{\kappa_k\gamma\cdot Pr(y|\alpha,\beta,N_{kt})}{1+\rho}\underbrace{\left(E[V_k(\phi',\alpha+y,\beta+N_{kt}-y,t+1)]-U(\alpha+y,\beta+N_{kt}-y,t+1)\right)}_{\text{surplus from involuntary J2J within k}}\\
&+\sum_{y=0}^{N_{kt}}\frac{\gamma\cdot Pr(y|\alpha,\beta,N_{kt})}{1+\rho}\sum_{k'\neq k}\Lambda_{kk'}(\pi(y),t+1)\underbrace{\left(E[V_{k'}(\phi',\alpha+y,\beta+N_{kt}-y,t+1)]-U(\alpha+y,\beta+N_{kt}-y,t+1)\right)}_{\text{surplus from moving to another group k'}}\\
&+\sum_{y=0}^{N_{kt}}\frac{\gamma\cdot Pr(y|\alpha,\beta,N_{kt})\,\Lambda_{kk}(\pi(y),t+1)}{1+\rho}\underbrace{\int_{\phi}^{\overline{\phi}_k}\left(V_k(\phi',\alpha+y,\beta+N_{kt}-y,t+1)-V_k(\phi,\alpha+y,\beta+N_{kt}-y,t+1)\right)dF_k(\phi')}_{\text{surplus from moving up within k}}\\
&+\sum_{y=0}^{N_{kt}} Pr(y|\alpha,\beta,N_{kt})\cdot\frac{1-\mu_t-\delta-\kappa_k-\sum_{k'\neq k}\Lambda_{kk'}(\pi(y),t+1)}{1+\rho}\cdot\underbrace{V_k(\phi,\alpha+y,\beta+N_{kt}-y,t+1)}_{\text{staying}}
\end{aligned}
$$

where posterior mean $\pi(y) \coloneqq \dfrac{\alpha+y}{\alpha+\beta+N_{kt}}$.

Workers employed in k with public state (α,β,t) produce signal y with probability $Pr(y|\alpha,\beta,N_{kt}) = \binom{N_{kt}}{y}\frac{B(\alpha+y,\beta+N_{kt}-y)}{B(\alpha,\beta)}$. The current output flow $m_k(\phi,\theta,t) = \phi\cdot(1+E[Y|\theta,N_{kt}]\cdot\eta)$ varies by employer group due to the heterogeneity in the opportunities to publish N_{kt}. The posterior belief $\text{Beta}(\alpha+y,\beta+N_{kt}-y)$ varies by k and t for the same reason. At $t=10$, the public state no longer evolves: production continues but signals no longer update the belief, so the continuation values and arrival rates on the right-hand side of (A.16) are evaluated at the current state $(\alpha,\beta,10)$, and $\mu_{10}=\overline{\mu}$.

The main change relative to joint value (3) in the benchmark model is the introduction of cross-group moves at rates $\Lambda_{kk'}(\pi,t)$ for $k'\neq k$. I assume workers cannot bargain across groups, so a worker hit by a cross-group shock compares the new offer against unemployment rather than her current match. Under this assumption, the worker always accepts the cross-group offer.[46] This keeps the model tractable by avoiding direct comparisons of match values across groups, which would involve non-wage amenities and idiosyncratic

[46] Within-group transitions remain efficient: workers accept offers from more productive firms within the same group. The argument of Lemma 1 applies group by group: the cross-group terms in (A.16) are independent of the current ϕ and drop out of the derivative, whose within-group coefficient $1-\mu_t-\delta-\kappa_k-\sum_{k'\neq k}\Lambda_{kk'}(\pi(y),t+1)-\gamma\Lambda_{kk}(\pi(y),t+1)\overline{F}_k(\phi)$ is nonnegative when the shock probabilities sum to at most one at every posterior.

preferences over employers (Card, Cardoso, Heining, and Kline 2018; Lentz et al. 2026).

D.2 Flow-balance Equations in Steady State

Denote by $u(\theta,\alpha,\beta,t)$ the density of unemployed workers with ability θ and public state (α,β,t), and by $l_k(\theta,\alpha,\beta,t)$ the density of workers employed in group k. At labor market entry, all workers are employed. Entrants are characterized by priors (10) with shape parameters depending on initial characteristics and the initial placements. Since workers start employed, $u(\theta,\alpha,\beta,0)\equiv 0$.

For $t\in\{1,\ldots,9\}$, the flow-balance equation for unemployed workers equates the outflow (workers exiting permanently and workers receiving job offers) with the inflow from layoffs:

$$\begin{aligned}&(\mu_t+\sum_{k\in\{A,I,T\}}\Lambda_{0k})\cdot u(\theta,\alpha,\beta,t) \\ &=\delta\cdot\sum_{k\in\{A,I,T\}}\sum_{y=0}^{N_{k(t-1)}} Pr(y|\theta,N_{k(t-1)})\, l_k(\theta,\alpha-y,\beta-(N_{k(t-1)}-y),t-1).\end{aligned} \tag{A.17}$$

Since the public state (belief and experience) is held fixed once workers accumulate ten years of experience, the flow-balance equation for the unemployed at $t=10$ is:

$$\begin{aligned}&(\overline{\mu}+\sum_{k\in\{A,I,T\}}\Lambda_{0k})\cdot u(\theta,\alpha,\beta,10) \\ =&\delta\cdot\sum_{k\in\{A,I,T\}}\sum_{y=0}^{N_{k9}} Pr(y|\theta,N_{k9})\, l_k(\theta,\alpha-y,\beta-(N_{k9}-y),9)+\sum_{k\in\{A,I,T\}}\delta\cdot l_k(\theta,\alpha,\beta,10).\end{aligned} \tag{A.18}$$

Unlike (A.17), the right-hand side also includes inflows from workers already in the absorbing state $(\theta,\alpha,\beta,10)$, since both belief and experience remain fixed after $t=10$.

For workers employed at experience $t\in\{1,\ldots,9\}$, the outflow from group k of workers with ability θ, experience t, and employer belief $\text{Beta}(\alpha,\beta)$ is the density $l_k(\theta,\alpha,\beta,t)$, since experience increments from t to $t+1$ by the end of the period. The inflow comes from three sources: hires from unemployment, stayers in k whose belief is updated to $\text{Beta}(\alpha,\beta)$,

and movers from other groups $k' \neq k$.

$$
\begin{aligned}
l_k(\theta,\alpha,\beta,t) = & \underbrace{\Lambda_{0k} \cdot u(\theta,\alpha,\beta,t)}_{\text{from unemployment}} \\
& + \underbrace{(1-\mu_t-\delta-\sum_{k'\neq k}\Lambda_{kk'}(\pi,t)) \sum_{y=0}^{N_{k(t-1)}} Pr(y|\theta,N_{k(t-1)}) \cdot l_k(\theta,\alpha-y,\beta-(N_{k(t-1)}-y),t-1)}_{\text{stayers in k}} \\
& + \underbrace{\sum_{k'\neq k}\Lambda_{k'k}(\pi,t) \sum_{y=0}^{N_{k'(t-1)}} Pr(y|\theta,N_{k'(t-1)}) \cdot l_{k'}(\theta,\alpha-y,\beta-(N_{k'(t-1)}-y),t-1)}_{\text{from other } k'\neq k}
\end{aligned} \tag{A.19}
$$

At $t = 10$, the outflow comprises workers who exit the labor market permanently, workers laid off into unemployment, and workers who transition to other groups $k' \neq k$. The inflow, in addition to the three sources on the right-hand side of (A.19), includes workers hired from other groups in the absorbing state.

$$
\begin{aligned}
& (\overline{\mu}+\delta+\sum_{k'\neq k}\Lambda_{kk'}(\pi,10)) \cdot l_k(\theta,\alpha,\beta,10) \\
= & \underbrace{\Lambda_{0k} \cdot u(\theta,\alpha,\beta,10)}_{\text{from unemployment}} + \underbrace{(1-\overline{\mu}-\delta-\sum_{k'\neq k}\Lambda_{kk'}(\pi,10)) \sum_{y=0}^{N_{k9}} Pr(y|\theta,N_{k9}) \cdot l_k(\theta,\alpha-y,\beta-(N_{k9}-y),9)}_{\text{stayers in } k \text{ from } t=9} \\
& + \underbrace{\sum_{k'\neq k}\Lambda_{k'k}(\pi,10) \sum_{y=0}^{N_{k'9}} Pr(y|\theta,N_{k'9}) \cdot l_{k'}(\theta,\alpha-y,\beta-(N_{k'9}-y),9)}_{\text{from other groups at } t=9} + \underbrace{\sum_{k'\neq k}\Lambda_{k'k}(\pi,10) \cdot l_{k'}(\theta,\alpha,\beta,10)}_{\text{from other groups in absorbing state}}
\end{aligned} \tag{A.20}
$$

At top firms ($k = T$), whether the share of workers at top firms increases with posterior mean $\pi = \frac{\alpha}{\alpha+\beta}$ depends on whether the cross-group inflow rates Λ_{IT} from industry non-top firms and Λ_{AT} from academia rise more steeply with π than the outflow rates from top firms — Λ_{TI} and Λ_{TA}. Parameters in $\Lambda_{kk'}(\pi,t)$ are identified from transitions between groups and the response of these transitions to publications (see Table A.1). The semi-elasticities of moving to top firms with respect to posterior mean are strongly positive for workers at non-top firms and for postdocs in academia (Table A.3).

Appendix E. Estimation Details

Empirical Moments and Bootstrapped Variance

Table A.1 reports the empirical moments from the training data. I run 1,000 bootstrap replications that resample workers with replacement and report the bootstrapped standard error of each moment. Denote by $\widehat{\Sigma}$ the bootstrapped variance–covariance matrix of the empirical moments.

Weight Matrix

The SMM objective is the weighted distance between the empirical and simulated moments. I weight the moments by a shrinkage matrix Ω_ω that blends the diagonal of $\widehat{\Sigma}$ with the identity matrix:

$$\Sigma_\omega := \omega \cdot \frac{K}{\operatorname{tr}(\widehat{\Sigma})} \operatorname{diag}(\widehat{\Sigma}) + (1-\omega) \cdot I, \text{ and } \Omega_\omega := \left[\frac{\operatorname{tr}\left(\Sigma_\omega^{-1}\right)}{K} \cdot \Sigma_\omega \right]^{-1} \tag{A.21}$$

where K is the number of moments in each step. The diagonal component is rescaled to have trace K, matching the identity, so that the two components are on the same scale; the second line normalizes the weight matrix so that $\operatorname{tr}(\Omega_\omega) = K$. The shrinkage toward the identity prevents moments with very small bootstrapped variances from dominating the objective. In practice, I never invert Σ_ω explicitly: I factor the bracketed matrix as LL' via the Cholesky decomposition and evaluate the objective $g'\Omega_\omega g$, where g is the gap between simulated and empirical moments, by solving two triangular systems — faster and more stable numerically. The shrinkage weight ω is selected from $\{0, 0.1, 0.3, 0.5, 0.7, 0.9\}$ based on model fit in the holdout sample in both steps.

Simulation

In the first step, each simulation initializes 25 cohorts of 1,250 labor market entrants by bootstrapping workers at $t = 0$ with replacement from the training data. The second step simulates the full panel for 50 periods, with a new entrant cohort in each period, and drops the first 25 periods as burn-in. All simulation draws are held fixed across parameter evaluations: true ability θ is drawn by applying the inverse CDF of $\text{Beta}(\alpha(X,k), \beta(X,k))$

to fixed uniform draws, publication outcomes are generated from fixed pools of uniform draws, and the bootstrap row indices for entrant cohorts are fixed.

Initial Guesses and Refinement

Initial guesses are generated from quasi-random (Sobol) draws over the full parameter box, together with local Gaussian perturbations of varying magnitudes around a baseline guess. In the first step, I screen all 285 candidate triples $\vec{N}_0 = (N_{A0}, N_{I0}, N_{T0})$ from a single initial guess of the prior parameters, keep the top 15 triples, and re-estimate the prior parameters under each triple from 24 initial guesses (360 combinations in total). The best 45 combinations of $\vec{N}_0$ and estimated prior parameters — across triples and initial guesses — are carried to the second step. In the second step, I run one iteration of the optimization from 23 initial guesses of the remaining 28 parameters for each candidate carried over. I then refine the search from the top 30 full parameter vectors, minimizing the distance in steps 1 and 2 sequentially, and report the estimates that achieve the best overall fit in the holdout sample in Table A.2 and A.3.

Counterfactual market beliefs

Let $X_i = (1, W_i, \text{acad}_i, \text{top}_i)$ collect worker i's initial characteristics and initial placement. The estimated model specifies ability as

$$\theta_i \mid X_i \sim Beta\big(\alpha(X_i), \beta(X_i)\big), \qquad \alpha(X) = e^{a'X}, \quad \beta(X) = e^{b'X},$$

and in the benchmark the market prior coincides with this distribution. Each counterfactual specifies an information set: the observed entries X_O are the components of W the market retains, plus $(\text{acad}, \text{top})$ if initial sorting is on; the remaining entries X_U are unobserved. The distribution of true ability is unchanged in very counterfactual; only the market's prior changes.

Exact belief. Given only $X_O = x_O$, the Bayes-consistent prior is the conditional mixture

$$p(\theta \mid x_O) = \int Beta\big(\theta; \alpha(x_O, x_U), \beta(x_O, x_U)\big)\, dF(x_U \mid x_O),$$

with first two moments, by iterated expectations,

$$m(x_O) = E\left[\frac{\alpha(X)}{\alpha(X)+\beta(X)}\,\middle|\,x_O\right], \qquad E[\theta^2 \mid x_O] = E\left[\frac{\alpha(X)(\alpha(X)+1)}{(\alpha(X)+\beta(X))(\alpha(X)+\beta(X)+1)}\,\middle|\,x_O\right],$$

and $v(x_O) = E[\theta^2 \mid x_O] - m(x_O)^2$.

Empirical implementation. I replace $F(x_U \mid x_O)$ with the empirical distribution within cells of the estimation sample at $t = 0$: cell $c(i)$ collects workers matching i exactly on observed binary characteristics and on quantile bins of observed continuous ones. With n_c workers in cell c,

$$\hat{m}_c = \frac{1}{n_c}\sum_{i\in c}\frac{\alpha_i}{\alpha_i+\beta_i}, \qquad \hat{v}_c = \frac{1}{n_c}\sum_{i\in c}\frac{\alpha_i(\alpha_i+1)}{(\alpha_i+\beta_i)(\alpha_i+\beta_i+1)} - \hat{m}_c^2,$$

where $(\alpha_i,\beta_i) = \big(\alpha(X_i),\beta(X_i)\big)$. Because the mixture is not a Beta distribution, I assign the market the Beta prior matching its mean and variance:

$$\bar{s}_c = \frac{\hat{m}_c(1-\hat{m}_c)}{\hat{v}_c} - 1, \qquad a_E = \hat{m}_c\,\bar{s}_c, \qquad b_E = (1-\hat{m}_c)\,\bar{s}_c.$$

Every entrant drawn from cell c receives prior $Beta(a_E, b_E)$, updated conjugately by subsequent publications. When initial sorting is off, the first job is drawn independently of X_i from the empirical placement distribution, and $(\text{acad}, \text{top})$ are excluded from X_O, since a random placement is uninformative.

Belief precision. Within each cell, the mean of the prior is preserved. By the law of total variance, $\hat{v}_c = E[Var(\theta \mid X) \mid c] + Var\big(m(X) \mid c\big)$: hiding characteristics converts the between-worker dispersion in expected ability they explained into additional prior uncertainty, so the belief precision $\bar{s}_c$ lies below the full-information precision. Aggregating, $E_c[\hat{v}_c] = E_X[Var(\theta \mid X)] + E_c[Var(m(X) \mid c)]$, strictly exceeding the average full-information variance whenever hidden characteristics predict ability within cells.

Appendix References

BAGGER, JESPER AND RASMUS LENTZ (2019): "An Empirical Model of Wage Dispersion with Sorting," *The Review of Economic Studies*, 86 (1), 153–190.

CARD, DAVID, ANA RUTE CARDOSO, JOERG HEINING, AND PATRICK KLINE (2018):

"Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36 (S1), S13–S70.

GIBBONS, ROBERT AND LAWRENCE KATZ (1992): "Does Unmeasured Ability Explain Inter-Industry Wage Differentials," *The Review of Economic Studies*, 59 (3), 515–535.

GIBBONS, ROBERT AND MICHAEL WALDMAN (1999): "A Theory of Wage and Promotion Dynamics Inside Firms," *The Quarterly Journal of Economics*, 114 (4), 1321–1358.

LENTZ, RASMUS (2010): "Sorting by search intensity," *Journal of Economic Theory*, 145 (4), 1436–1452, search Theory and Applications.

LENTZ, RASMUS, JONAS MAIBOM, AND ESPEN R. MOEN (2026): "Directedness in Search," (26022).

MARCO, ALAN C., AMANDA MYERS, STUART J.H. GRAHAM, PAUL D'AGOSTINO, AND KIRSTEN APPLE (2015): "The USPTO Patent Assignment Dataset: Descriptions and Analysis," USPTO Economic Working Paper 2015-2, United States Patent and Trademark Office.

MARX, MATT AND EMMA SCHARFMANN (2024): "Does Patenting Promote the Progress of Science? Evidence from Patent-Paper Pairs," Working Paper.

MILLER, RICHARD D. (2020): "Technical Documentation for the 2019 Patent Examination Research Dataset (PatEx) Release," USPTO Economic Working Paper 2020-4, United States Patent and Trademark Office, Office of the Chief Economist.

RUDIN, WALTER (1976): *Principles of Mathematical Analysis*, New York: McGraw-Hill, 3rd ed.

STRASSEN, VOLKER (1965): "The Existence of Probability Measures with Given Marginals," *The Annals of Mathematical Statistics*, 36 (2), 423–439.