Reveal or Conceal?

Employer Learning in the Labor Market for Computer Scientists

Alice H. Wu*

November 1, 2023

Job Market Paper - Click here for the latest version.

Abstract

The efficient allocation of labor relies on the identification of talent. When employee output is not publicly observable, employers have an incentive to take advantage of private information, potentially leading to the misallocation of labor among firms. This paper provides empirical evidence of employer learning and quantifies the impact of learning on job mobility and innovation outputs in the labor market for computer science (CS) Ph.D.'s. CS conference proceedings provide public information on research effort by existing CS workers. Among papers authored by researchers from industry, about one-quarter can be matched to a contemporaneous patent application - an indicator of a more valuable innovation. Yet the fact of the application remains private information at the incumbent employer for 18 months. Consistent with public learning, researchers with a new paper have higher inter-firm mobility rates than do coworkers without a paper. Initially, authors of papers with a matched patent are *less* likely to move than authors without a patent application. But once the patent application becomes public, their mobility rates cross over. Authors of papers with a matched patent are also 35% more likely to move to a top tech firm. These patterns confirm the predictions of a model in which incumbent firms have initially private information on more productive researchers. Structural estimates of the model suggest that if papers and patents were disclosed simultaneously, high-ability workers would sort more quickly to high-productivity firms. The implied increase in allocative efficiency would increase innovation outputs by about 5%.

1 Introduction

The identification of talent is critical to the efficient allocation of labor in the economy. A large body of existing research suggests that worker abilities are only partially revealed prior to their labor market entry, and that substantial learning by employers occurs over the first decade or so of work (e.g., Altonji and Pierret 2001; Farber and Gibbons 1996; Pallais, 2014; Tervio, 2009). Thus far, most empirical research has focused on "public learning" – information that is assumed to be revealed simultaneously to both actual and potential employers. Much less is known about the revelation of information that an incumbent employer can see but that other potential employers only learn later.

How are workers affected when an incumbent employer has better information about their ability than other potential employers? Existing theoretical work gives ambiguous guidance. On the one hand, an incumbent employer may hide private information by delaying promotions, slowing the career growth of workers (Waldman 1984). On the other hand, privately informed employers have more incentive to invest in workers, potentially increasing future productivity (e.g., Acemoglu and Pischke 1998; Ferreira and Nikolowa 2022; Strobl and Van Wesep 2013). Different answers from the alternative models highlight the need for a unified framework that incorporates employers' incentives to enhance productivity while retaining talent.

In this paper I offer both theoretical and empirical answers to this important question, focusing on the labor market for highly trained computer scientists. Every year about 4,000 Ph.D.'s graduate in computer science (CS) or closely related fields in the United States.[1] These researchers often continue to publish at academic conferences, yielding public information on their research ability. In this market I can also measure private information held by incumbent employers. About 24% of papers from the industry are accompanied by a patent application, the existence of which can remain private information with the incumbent employer for 18 months.[2] I build this institutional feature into a dynamic model of employer learning with mobility frictions, where firms set wages and allocate workers between innovation and routine tasks based on the information they can observe. The model's predictions guide my reduced-form tests for asymmetric learning based on the delayed disclosure of a patent matched to a paper.

Using web-scraped public LinkedIn profiles of CS Ph.D.'s matched with their dissertations and conference papers, I find that workers with a new paper are more likely to move between employers within a year than similar coworkers. The job mobility responses to a paper are strongest among those working outside the top firms, which I define in this paper as Google (Alphabet), Facebook

---

[1] There are roughly three times as many new Ph.D.'s in CS/EE /Info Science combined as in Economics, according to the NSF Survey of Earned Doctorates (2021).

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

(Meta), Amazon, Microsoft, IBM, and Apple. I interpret these findings as evidence of public employer learning. There is also evidence of asymmetric learning. Before the public revelation of a patent, workers with a new paper and matched patent experience fewer moves than do coworkers with a paper only. After the revelation, however, they are 13% more likely to move out of a non-top firm and 35% more likely to move to a top firm. According to the structural model estimates, removing the delayed disclosure (equivalent to increasing public information) would increase the annual upward mobility rate of workers with a patent from non-top to top firms by 32% (comparable with the reduced-form estimate). Total innovation outputs by computer scientists would increase by about 5%, driven by an increase in positive assortative matching and the fact that employers can allocate new recruits more efficiently given better information.

The main analysis begins with a dynamic model of employer learning. Firms vary in productivity in innovation versus other (more routine) tasks, and workers vary in research ability that matters for innovation only and is unknown to everyone initially. Bayesian firms update beliefs about workers based on their outputs in innovation tasks. There is a key trade-off between learning and retention stressed in the model: allocating workers to more innovation tasks helps an incumbent employer identify high-ability workers faster and improve productive efficiency in the future, but it also increases the risk that high-ability workers will be recognized and poached by outside employers. This trade-off is particularly important in imperfectly competitive labor markets where firms have some wage-setting power.

Building on Card, Cardoso, Heining, and Kline (2018), I allow workers to have idiosyncratic preferences over firms. Conditional on information about workers, firms set wages just like how they would set prices in oligopoly (e.g., Bresnahan 1981; Berry, Levinsohn, and Pakes 1995; Goldberg 1995). In equilibrium, wages are posted simultaneously and are set by firms to maximize their own profits, taking as given the wages at other firms. Employers also consider turnover risks when allocating workers to talent-revealing innovation tasks. Less productive firms set lower wages, face higher turnover, and allocate fewer innovation tasks in equilibrium.

The model generates three testable predictions that link information revelation to job mobility and future productivity: (1) Workers with newly revealed innovation are more likely to move between firms and move to more productive firms than are similar workers without such signals. (2) Job mobility is suppressed for workers with positive signals that are observed by the incumbent employer but unknown to potential employers outside. (3) Workers who are privately known to be higher ability are more productive when they stay with their incumbent employers, which allocate labor more efficiently given the superior information.

The labor market for computer scientists offers an interesting and policy-relevant setting to test for employer learning and quantify the impacts of asymmetric information on labor market and

2

innovation outcomes. The majority of new CS Ph.D.'s in recent years enter the private sector. The share of conference proceedings with an author from industry has been rising over the past decade. Those papers represent on-the-job research that otherwise would have been hard to observe, and that matters for recruiting research scientists and those in similar roles.[3] Workers with publications post Ph.D. are increasingly sorted into top firms in industry (Figure 1).

The data that I employ was assembled as follows. Using the ProQuest dissertation database and school-specific sources, I assembled a list of around 96,000 Ph.D. graduates from the top 60 CS departments in the United States between 1980 and 2021. I then searched LinkedIn and found more than 40,000 public profiles that can be matched to a dissertations by full name, school, and year of graduation, and that have post-Ph.D. full-time employment records. These researchers' profiles are then matched to publications from 80 conferences and two machine learning journals, as well as to patent applications. It allowed me to keep track of a person's research outputs during each spell of employment.

About 24% of publications from authors in my database with an industry affiliation at the time of the paper can be matched to a patent application filed around the same time (vs. 5% in academia). Table 1 provides a few examples. Papers that are matched to a patent are higher quality on average, receiving more citations from other research papers (excluding self-citations). People who produce such papers are also more likely to be end up working at a top tech firm (Figure 1). Both facts suggest firms are able to identify valuable innovations and claim exclusive rights to the inventions early on. The fact of a patent application remains private information for 18 months by default (AIPA 1999).[4] This institutional feature allows me to observe a margin of asymmetric learning at which the incumbent employer (as the assignee) has full knowledge of the ongoing patent application while the outside market observes only the research paper.

I test the model's predictions by regressing job mobility outcomes on whether a worker has new papers and matched patents, and whether her previous papers are matched to patents that have recently been made public. The reduced-form strategy compares inter-firm mobility patterns of workers with a new paper versus similar coworkers who did not. The baseline specification controls for worker education and job experience since PhD, as well as firm-year fixed effects to absorb firm-specific shocks to mobility such as a layoff.

The mobility responses to new research signals are the largest and most statistically significant for workers at non-top firms, defined as industry employers that are not among the top. I refer to

---

[3]For example, see job postings in Appendix Figure B1.

[4]There are a number of exceptions to the 18-month publication rule under 35 U.S.C. 122(b) of AIPA (1999). Applications that are no longer pending (e.g., a patent is issued) are published immediately. Upon filing, applicants may submit a non-publication request such that the applications remain private until they are no longer pending (U.S.C.122(b)-2A-B). The non-publication request must be rescinded if the invention has been or will be the subject of an application filed in another country.

workers without a paper as the baseline $(0, 0)$ group, workers with a paper but no matched patent as $(1, 0)$, and those with a matched patent as $(1, 1)$. Relative to the $(0, 0)$ group, the $(1, 0)$ group at non-top firms is 3.5 percentage points (henceforth ppt's) or 27% more likely to move between employers in the next year, and 1.8 ppt's (51%) more likely to move to a top firm. Importantly, they are also more mobile than the $(1, 1)$ group in the year immediately after the paper appears. This is consistent with the second prediction of my model: incumbent employers with private information should be more likely to retain workers who are identified as more productive (similar to Greenwald, 1986).[5] Thus, researchers with a new paper that is matched to a patent have exit rates from their incumbent employer that is lower than the rate for researchers with a new paper but no patent.

Once patent applications that are matched to papers become public information, workers with such delayed signals become more mobile than similar coworkers. Looking out three years, a worker with a new paper and a matched patent application at non-top firms has a 1.8 ppt (13%) higher mobility rate than one with a new paper alone, and 1.3 ppt (35%) higher probability of moving to a top tech firm. In contrast, papers without a patent have only a short term impact on mobility outside academia, conditional on new research outputs and other observables.

The evidence of asymmetric learning, from the delayed mobility for authors of papers with a matched patent, is robust if I control for patent applications that are not matched to any paper. This suggests that the mobility patterns at the margin of asymmetric learning are not driven by the act of patenting itself, but rather by the different information about talent in papers with versus without a matched patent. The results also hold if I exploit within-person variation in research outputs. Workers with new papers or newly revealed patents are also more likely to move into higher-wage firms, get promoted, and become a scientist rather than engineer. Together, these findings confirm the model predictions that labor market mobility increases in public information, and that mobility is delayed when an incumbent employer has superior information earlier than the market.

One concern with my approach of comparing mobility patterns of workers with versus without patents is that workers may want to stay with an incumbent firm to receive credit for their work. However, U.S. patent law requires an application to acknowledge all contributors: the patent application would be annulled if the name of an inventor was removed. I also show that being an inventor of a *granted* patent, which typically occurs three or more years after the application, does not affect job mobility conditional on more recent research. This is consistent with evidence in Kline, Petkova, Williams, and Zidar (2019) that wages of most employees (not founders or managers) are unaffected by granted patents.

Another concern is that the existence of a patent application for work associated with a new paper may be disclosed in other ways before the official publication by the patent office. To evaluate

---

[5]A difference from the model in Greenwald (1986) is that $(1, 0)$ or $(0, 0)$ workers still receive a positive wage because they are productive in routine tasks.

4

this concern, I selected a random sample of papers and found no mention of (or citation to) a matched patent application that is yet to be disclosed. Moreover, non-disclosure clauses can make it risky for workers to disclose unpublished patents.[6] Workers may showcase their patents under the "Accomplishments" section on a LinkedIn profile. However, in my data nearly all patents mentioned in this section are granted.[7]

An important question that remains is how much asymmetric information matters for innovation. As predicted by the model, the $(1,1)$ group produces significantly more papers, especially higher-quality papers with a matched patent, than the $(1,0)$ group in the following three years. But the gap is larger among stayers, than among movers who enter a new employer that cannot tell a $(1,1)$ apart from a $(1,0)$ immediately. Incumbent employers with superior information let $(1,1)$ workers spend more time on innovation tasks earlier than would a less informed but equally productive new employer. This productivity difference also reflects the adverse selection of movers under asymmetric information (e.g., Greenwald 1986).[8]

The impacts of asymmetric learning on total innovation remain ambiguous in the reduced-form analysis. $(1,1)$ workers could have moved to more productive employers earlier and devoted even more time to innovation. To resolve the ambiguity, I estimate the model to maximize the joint likelihood of workers' movements between 16 groups of employers, and their innovation outputs in a balanced panel that covers the first decade of post-Ph.D. career. Matched patent applications are assumed to be revealed one year later than the original papers. Initially, it is highly uncertain if a worker will be productive in research. But model estimates suggest that employers update beliefs according to the observable innovation signals, the majority of high-ability researchers can be idenitified in ten years. Workers who are perceived to be productive are increasingly concentrated at top firms or in academia, matching the patterns in data (e.g., Figure 1).

To quantify the impact of asymmetric learning, I compare the equilibrium outcomes under asymmetric information versus a counterfactual scenario where matched patents are disclosed simultaneously as papers. Given the maximum-likelihood estimates from the asymmetric benchmark, I solve for each employer's optimal wages and task allocations under symmetric disclosure. Simulations of the career paths of CS workers under these two information schemes show that the rates at which workers produce a paper or a paper with a matched patent would increase by 5-6% if the information were more symmetric between employers. On average $(1,1)$ workers are 13% more

---

[6]For example, *Cloudera Inc. v. Databricks Inc. et al* cites a nondisclosure clause from the Proprietary Information and Inventions Agreement signed by a former employee: "[I] will hold in confidence and not disclose or, except within the scope of my employment, use any proprietary information", which includes inventions developed on the job.

[7]In recent years there are almost zero mentions of patent applications in the same year a worker has a new CS paper (Appendix Figure B5).

[8]Similar to Greenwald (1986), incumbent employers pay a higher wage to privately known $(1,1)$ than $(1,0)$. Conditional on searching for new jobs, $(1,1)$ workers have a higher probability of choosing the incumbent employer again, given the higher wage it offers. As a result, $(1,1)$ are less likely to move than $(1,0)$ under asymmetric information.

likely to move from a non-top firm, and 32% more likely to move from non-top to top firms than $(1,0)$ who produce a paper only. The increases in job mobility among $(1,1)$ workers who produce a high-quality paper are comparable to the regression estimates in the full sample. $(1,1)$ workers who move to a new employer would also spend more time on innovation tasks than before, whereas $(1,0)$ movers spend less time on innovation. The increase in positive assortative matching and more efficient task allocation among new recruits can explain the 5% increase in total innovation outputs.

In summary, this paper provides novel evidence of asymmetric employer learning in a high-skilled, innovation-intensive labor market. The empirical tests for learning are derived from a dynamic game between employers, and exploit the institutional feature that matched patent applications are disclosed later than the papers themselves. The delay in the job mobility of workers who produce a high-quality paper with a matched patent, especially from non-top to top firms in industry, provides evidence for asymmetric learning. The counterfactual analysis further suggests innovation would increase if information disclosure were symmetric, providing a more complete picture of the impacts of asymmetric learning.

1.1 Related Literature

This paper contributes to several strands of the literature. First, it provides new empirical evidence of employer learning by estimating mobility responses to signals about a worker's research ability. The canonical tests of learning rely on the differences between researchers' information about workers and the labor market's. The increasing correlation between wages and AFQT scores (observed by researchers but not firms) over time is often used as evidence of learning (Altonji and Pierret 2001; Farber and Gibbons 1996). The underlying model of these studies posits that employers update their belief when new signals arrive, but those signals are rarely observable except when personnel records are available (Kahn and Lange 2013). The large volume of CS papers in this setting allows me to observe public information on a worker's on-the-job research. Employers also explicitly emphasize publication records as an important qualification for research jobs. The delayed disclosure of a matched patent application further allows me to contribute to the empirical studies of asymmetric learning. I exploit the asymmetric information between an incumbent employer and the labor market directly, rather than the wedge between we researchers' and firms' information (Schönberg 2007; Kahn 2013). The evidence of asymmetric learning in this paper supports the important insight in Schönberg (2007) that learning is more asymmetric when the employees involved are higher skilled.

Second, this paper contributes to the theoretical literature on asymmetric learning. The classic learning models often begin with homogeneous players (employers) in a perfectly competitive (labor) market, which are reasonable simplifying assumptions to focus on the implications of asymmetric

6

information (Boozer 1994; Greenwald 1986; Henricks and Porter 1988; Li 2012). Relaxing the homogeneity and perfect competition assumptions generates a richer set of predictions and gets closer to the CS labor market in reality. Allowing firms to have some labor market power introduces a crucial trade-off between learning and retention. Employers are willing to invest in learning when workers who are learned to be high-ability do not move away immediately. This result is closely related to the discussions of general skill training in a monopsonistic market (Acemoglu and Pischke 1998; Manning 2003; Stevens 1994). Adding heterogeneity between firms, as in Gibbons and Katz (1992), enriches the predictions on job mobility. On average movers are adversely selected (Gibbons and Katz 1991; Greenwald 1986). But there is assortative matching between workers with higher research ability and more innovative firms where workers can be extra productive. Further, less productive firms face higher turnovers in equilibrium and allocate less innovation tasks. The contrast between low and high-productivity firms is related to the multiple equilibria in Acemoglu and Pischke (1998).

This paper contributes to the innovation literature by linking research papers and patents in computer science. To the best of my knowledge, this is the first large-scale effort to match papers with patent applications in computer science, which relies on not only the text similarity in titles and abstracts, but detailed information about the authors and their employment history. A similar exercise has been done in biotechnology, where Magerman, Van Looy, and Debackere (2015) find about 600 paper-patent pairs based on text and requiring at least one author in common. Ahmadpoor and Jones (2017) look more broadly across disciplines and establish paper-patent linkage from citations. They find that research works at the "dual frontier", comprising papers that cite a patent and patents that cite a paper, are more impactful than works away from this citation frontier. Despite the differences in matching, their finding is similar to what I observe in computer science that CS papers with a matched patent application receive more citations than papers without a patent.

Last but not least, this paper is related to a growing literature on the labor market power of employers in tech. Tech companies are known to exploit their monopsony power, through non-compete contracts and collusive no-poaching (no cold-call) agreements (US Department of Justice 2010; Gibson 2023). This paper zooms into asymmetric information between employers as another source of monopsony power.

The remainder of this paper is structured as follows. Section 2 describes the CS labor market. Section 3 presents the dynamic model of employer learning and derives testable predictions of employer learning on job mobility and productivity. Section 4 outlines the data. Section 5 presents the empirical tests of asymmetric employer learning. Section 6 estimates the model and quantifies the impact of asymmetric learning on total innovation and talent revelation. Section 7 discusses future research directions.

2 Labor Market for Computer Scientists

The empirical setting that I study is the labor market for computer scientists. I will address how employers know whom to hire among CS Ph.D's, what tasks to allocate to them to advance innovation, and how employers can, at the same time, retain talent. I first describe three facts about the labor market for Ph.D. computer scientists to motivate the assumptions of the dynamic model of employer learning (Section 3).

Fact 1 (Industry Jobs Post Ph.D.) *The majority of new CS Ph.D. graduates and postdocs now enter the private sector, but they often continue to publish at academic conferences.*

Every year about 2,000 students graduate with a Ph.D. in Computer Science (CS) in the U.S., and 1,800 graduates with a Ph.D. in Electrical Engineering (NSF Survey of Earned Doctorates 2021).[9] Similar to economics, new CS Ph.D's may take a tenure-track or postdoc job in academia, or a job outside academia. The share of new CS Ph.D.'s who enter the industry as opposed to academia has been increasing over the past 20 years and exceeding 50% since 2017 (see Appendix Figure B2).

There are increasing opportunities to publish at academic conferences for computer scientists who are employed by the private sector. The share of conference proceedings with an author from the private sector was 35% in 2010 but 45% in 2022 (see Figure B3).[10]

Fact 2 (Research Papers and Patents) *Over 40% of the CS research papers that originate in the industry have a matched patent application filed around the same time. Papers with a patent receive more citations from other researchers in the future. Whether a paper has a matched patent is private information for more than a year after the patent is filed.*

Papers and patent applications are matched using criteria based on conditions for patentability as specified by patent laws - title 35 U.S.C. 102 (see Table 1 for examples, and Section 4 for details). Research from the industry is more than twice as likely to be matched to a patent application than research from academia. Papers with a patent, as shown in Figure 2, are higher-quality on average, receiving more citations in other research papers (excluding self-citations). Firms are able to identify valuable innovations and seek legal protection of inventions partially disclosed in a research paper.

Although the paper is public knowledge, whether a paper has a matched patent application or not is disclosed later. According to patent laws - title 35 U.S.C. 122 (AIPA 1999), patent applications remain private information for 18 months from the earliest filing date by default. During that

---

[9]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.

[10]The rise of research papers from the industry is largely driven by the recent advances in AI research that requires a considerable amount of computing power. In particular, in 2014 the winning model "AlexNet" of the Image contest popularized the use of GPUs in the training of deep learning models.

window, the incumbent employer (as the assignee on a patent) has full knowledge about the ongoing patent application while the outside market observes only the research paper. Workers themselves rarely advertise pending or unpublished patent applications (Appendix Figure B5).[11] A key feature of my work is the use of the delay in disclosing a patent to test for asymmetric employer learning.

Fact 3 (Assortative Matching in the CS Market) *Computer scientists who initially publish while working at non-top firms are more mobile than their coworkers, and are increasingly sorted into the top firms.*

About a quarter of CS papers from industry have an author from the top firms, i.e. {Google, Microsoft, IBM, Amazon, Facebook, Apple}. The productivity differences between firms generate a job ladder for computer scientists who are active in research. Figure 1 keeps track of the mobility into top firms by workers who start outside the top but produce different research signals. Workers who produce a research paper at non-top firms are twice as likely to show up at one of the top firms 10 years post Ph.D. as those without any paper. The upward mobility into a top firm further increases for workers who produce not only a paper but a matched patent application. These patterns suggest research outputs can help workers sort into more productive employers.

In summary, CS Ph.D.'s outside academia can publish papers, and these researchers become known entities in the field of CS and are more likely to move to top firms. A significant fraction of papers from the industry are filed as a patent application, which indicates higher-quality innovation but remains private information for more than a year. The delayed disclosure of a matched patent is built into the model as the source of asymmetric information between employers, motivating empirical tests for asymmetric learning and counterfactual analysis.

3   A Dynamic Model of Employer Learning

I develop a dynamic framework where firms recruit workers and allocate them to innovation tasks that facilitate learning about their research ability. Incumbent firms benefit from learning as they can allocate higher-ability researchers to more productive tasks. But they face a risk of increasing their employees' outside options to the extent their innovation outputs are public. Firms are more willing to learn when information about workers is less public, and when they have more monopsony power so that the learned talents are less likely to be poached. In the absence of information and labor market frictions, this framework is equivalent to the model of general skill training in Becker (1964). The benchmark model generates predictions regarding mobility between firms as information about workers is revealed, which I test in Section 5. It also provides a

---

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

structural framework to quantify the impact of employer learning on talent revelation and innovation productivity in Section 6.

3.1 Model Environment

I introduce the model environment that comprises heterogeneous workers and heterogeneous employers, and describe the labor market matching process. Time is discrete and finite in this framework.

3.1.1 Workers

Workers (indexed by i) are endowed with a binary one-dimensional research ability α_i, which can either be high H or low L. Upon labor market entry, there is information I_{i1} about research ability such as education, from which employers form a prior belief about ability $\pi_{i1} = Pr(\alpha_i = H|I_{i1})$. Workers and potential employers observe the same information and hold a common prior.

Each worker has 1 unit of time per period, and can split it between routine tasks and innovation tasks. H and L -ability workers are equally productive in routine tasks, but the H-ability are more productive in innovation.[12] Specifically, H-ability can produce an innovation, denoted by $y_{it} = 1$, with probability h per unit of time, whereas L-ability can do so with probability $l < h$.

Innovation outputs vary in quality. High-quality outputs increase the returns to innovation proportionally by $\theta > 0$. Conditional on any innovation ($y_{it} = 1$), the probability that it is to be high-quality ($\widetilde{y}_{it} = 1$) is $\widetilde{h}$ if it is produced by a H-ability worker, or $\widetilde{l}$ if it is produced by a L-ability worker. High-quality innovation helps further differentiate between H and L-ability when $\widetilde{h} > \widetilde{l}$. In summary, there are three potential outputs:

$$(y_{it}, \widetilde{y}_{it}) \in \{ \underbrace{(0,0)}_{\text{No Innov}} , \underbrace{(1,0)}_{\text{Low-quality}} , \underbrace{(1,1)}_{\text{High-quality}} \}$$

I define a function $q : [0,1] \rightarrow \mathbb{R}^+$ that yields the expected returns to innovation given an employer belief π that a worker is H-ability, per unit of time she spends on innovation:

$$q(\pi) = \underbrace{(\pi h + (1 - \pi)l)}_{\text{any innovation}} + \theta \times \underbrace{\left(\pi h \times \widetilde{h} + (1 - \pi)l \times \widetilde{l} \right)}_{\text{high-quality}}$$

---

[12]I could allow H-ability workers to also be more productive in routine tasks as long as the gap is larger in innovation tasks. That is, H has a comparative advantage in innovation tasks, similar to the setup in Gibbons and Katz (1992). Changing this assumption would not affect the model predictions in Section 3.3. To simplify matters, I focus on the case of equal H and L productivity in routine tasks.

3.1.2 Employers

Employers (indexed by j) are endowed with productivity f_j in routine tasks, and a proportional increase in productivity, g_j, in innovation tasks, both of which are public information. Employers simultaneously post wages ($w \in \mathbb{R}^+$) for a worker, based on their information about her at the beginning of each period. They also decide how much time the worker can spend on innovation tasks, $\tau \in [0, 1]$, if she becomes an employee. The total production at a firm each period is the sum of outputs across individual employees.[13] The marginal revenue product of a worker at j, as a function of belief π and allocation τ, can be written as:

$$MP_j(\pi, \tau) := f_j \times \left(\underbrace{(1 - \tau)}_{\text{Routine}} + \underbrace{g_j \times q(\pi) \times \tau}_{\text{Innovation}} - \underbrace{\zeta/2 \times \tau^2}_{\text{Cost}} \right) \tag{3.1}$$

where $f_j \times g_j$ represents j's productivity in innovation, and there is a convex cost of allocating workers to innovation tasks, determined by parameter ζ.[14] As in Gibbons and Katz (1992), there is matching between employers and workers. Research ability is valued more at an employer with greater g_j. I will show that in equilibrium the more innovative firms let workers spend more time on innovation tasks, making high-ability workers more productive than they would have been at lower-g_j firms. It is worth noting that less innovative firms can survive in this market: they may be more productive in routine activities, or provide amenities that are valued by some workers.

To capture the delayed disclosure of patents that are matched to papers in the CS labor market, I assume y_{it}, the indicator for any innovation such as a conference proceeding realized by the end of period t, becomes public information without further ado. The quality indicator $\widetilde{y}_{it}$ that represents whether a paper has a matched patent application, in contrast, is private information at incumbent employer $j(i, t)$ for a period. It will become public information at the beginning of $t + 2$ rather than $(t + 1)$. Employer learning is thus asymmetric under the delayed disclosure of quality information.

3.1.3 Labor Market Matching

Workers who are on the labor market observe the contracts posted by potential employers, which can be organized into four nests:

$$G(j) \in \{\text{Tenure-Track, Postdoc, Top Firms, Non-Top Firms}\}$$

---

[13]This model assumes away the joint production by workers. Identifying talent from team outputs can be difficult and warrants a more careful analysis in future work.

[14]This cost may include investment in computing power that often grows in a convex way as employees spend more time on innovation. It may also absorb the management costs of moving workers away from routine activities at a firm. For example, a firm may have to establish an in-house research lab, hire new managers, and establish a new performance evaluation system for workers who are increasingly involved in innovation tasks.

The first two nests represent academia, while the other two represent industry. Workers draw preferences across potential employers from a generalized extreme value distribution:

$$F(\{\epsilon_{itj}\}) = exp\left(-\sum_{G \in C}\left(\sum_{j \in G} exp(-\rho_G^{-1}\epsilon_{itj})\right)^{\rho_G}\right) \tag{3.2}$$

where C a worker's choice set at t.[15] The preferences are independent between nests and over time, but can be correlated within a nest if $\rho_G < 1$.

Given a wage offer w_{tj}, the utility of the worker choosing firm j is assumed to be:[16]

$$u_{itj} = b \times ln(w_{itj}) + \rho_{G(j)} \times \epsilon_{itj} \tag{3.3}$$

Assume $b \in (0, \infty)$ and $\forall G : \rho_G \in (0, 1]$ so that the labor market is imperfectly competitive.[17] Under the assumptions above, the labor supply by workers on the market is represented by the well-known nested logit model (McFadden 1973; Imbens and Wooldridge 2007).

All workers are on the labor market at $t = 1$ (the first year post PhD). Following a dynamic extension of Card et al. (2018), at $t > 1$ workers from nest G may get on the market again and search for new jobs with probability:

$$\lambda_G(\pi) = \lambda_{0,G} + \lambda_{1,G} \times \pi \tag{3.4}$$

which can be specific to the original nest G and depend on market belief about the worker denoted by π.[18] This formulation is equivalent to each worker drawing a random search cost $z \overset{d}{\sim} \Phi$, and only search for new jobs if $z < \overline{z}$, where $\Phi(\overline{z}) = \lambda$.[19] If a worker is on the market, she redraws GEV-distributed preferences across potential employers, as specified in (3.2). Other workers who are not on the market stay put and hold fixed the preferences they have drawn before.

---

[15] At $t = 1$, the choice set C includes all nests and thus all employers for workers who enter the labor market for the first time. At $t > 1$, postdoc jobs are no longer available for those who are not postdocs at $t = 1$. For workers from academia, the choice set includes all other employers with probability Λ_{AJ}, or only tenure-track employers the rest of the time. Similarly, for workers from the industry, tenure-track employers are in the choice set with probability Λ_{JA}. The academic-industry jobs opening rates $(\Lambda_{AJ}, \Lambda_{JA})$ are structural parameters to be estimated.

[16] I present an extension where workers take into account their future option values on the market in Appendix A4. Under simplifying assumptions, the option value of a worker is reduced to a preference for allocation to innovation tasks, τ, that enters her utility of choosing the employer today. The preference would be stronger for higher-π workers, and in equilibrium firms that offer higher τ set a relatively lower wage than in the benchmark model. That is, higher-π workers pay to do research early in their career, as in Stern (2004). Testable predictions on mobility in Section 3.3 remain the same in this extension.

[17] The labor supply of workers on the market is perfectly elastic over wages if $b \to \infty$ or $\rho_G \to 0$. I consider perfect competitive labor markets as special cases in Section 3.2.3.

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

[19] The λ's may also be interpreted as job arrival rates in the search models (e.g., Burdett and Mortensen 1998; Postel-Vinay and Robin 2002). But note in this discrete-time framework, workers who are on the market can see offers posted by all potential employers. Search models, in contrast, are often in continuous time and consider a single job arrival at any given time. Another difference from Postel-Vinay and Robin (2002), for example, is that there is no bargaining between a worker and incumbent employer when new offers arrive. Instead, workers redraw their preferences and choose a new employer (which may be the same as before).

3.1.4 Model Timeline & Information Structure

There are $T \geq 3$ discrete periods in this model. At least three periods are needed to fully capture the information revelation process: innovation is produced at an initial employer at $t = 1$; a paper may be published by the beginning of $t = 2$; whether a paper from $t = 1$ has a matched patent application is not revealed until $t = 3$.

1. ($t = 1$) Employers start with zero employees. All workers are on the labor market looking for jobs.
    (a) Given initial information $\{I_{i1}\}$ about workers, employers post wages $\{w_{i1j}\}$ simultaneously and choose the share of time each worker can spend on innovation tasks, $\tau_{i1j} \in [0, 1]$.
    (b) Each worker observes the wages posted by all firms and chooses an initial employer $j(i, 1)$ that maximizes her utility (3.3) at $t = 1$.
    (c) Innovation outputs are realized by the end of $t = 1$, $(y_{i1}, \widetilde{y}_{i1}) \in \{(0, 0), (1, 0), (1, 1)\}$.

2. ($t = 2$) Let I_{i2j} denote firm j's information about worker i at the beginning of $t = 2$.

$$\text{Incumbent } j(i, 1), \quad I_{i2j(i,1)} = I_{i1} \cup \{j(i, 1), y_{i1}, \widetilde{y}_{i1}\} \tag{3.5}$$
$$\text{Outside Firms } \forall j \neq j(i, 1), \quad I_{i2j} = I_{i1} \cup \{j(i, 1), y_{i1}\}$$

Information is symmetric if $y_{i1} = 0$, but asymmetric when $y_{i1} = 1$ and $\widetilde{y}_{i1}$ is unknown to outside employers.
    (a) Given info $\{I_{i2j}\}$, firms post new wages $\{w_{i2j}\}$ simultaneously and choose task allocation $\{\tau_{i2j}\}$, taking into account the expected labor supply from incumbent or outside employees.
    (b) Workers search for new jobs with probability (3.4), and choose a new employer that maximizes her utility (3.3). Other workers stay at the original employers.
    (c) Repeat 1(c).

3. ($t = 3$) Information is now summarized by $\{I_{i3j}\}$:

$$\text{Incumbent } j(i, 2) : I_{i3j(i,2)} = I_{i2j(i,2)} \cup \{\widetilde{y}_{i1}, y_{i2}, \widetilde{y}_{i2}\} \tag{3.6}$$
$$\text{Outside Firms } j \neq j(i, 2) : I_{i3j} = I_{i2j} \cup \{\widetilde{y}_{i1}, y_{i2}\}$$

The public information includes whether a paper from $t = 1$ has a matched patent application, and whether there is a new paper produced during $t = 2$. Repeat the rest of 2.

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

3.2 Backward Induction and Equilibrium

The simultaneous wage posting by employers can be viewed as multi-period Bertrand competition with incomplete and asymmetric information. Firms set wages for incumbent and new workers each period, taking the wages posted by other firms as given. The equilibrium concept is

subgame perfect Nash with wages as strategic variables, following the oligopoly solution concept in price-setting games (e.g., Bresnahan 1981; Goldberg 1995; Berry, Levinsohn and Pakes 1995). I state the problems of workers and firms, and derive the equilibrium backward.

3.2.1 Workers' Problem

Workers who are searching for jobs on the market solve the same problem each period. Given contracts $\{w_{itj}, \tau_{itj} : j \in C\}$ from potential employers in choice set C, worker i chooses her employer at t as follows:

$$j(i, t) = argmax_{j \in C} \ u_{itj} = b \times ln(w_{itj}) + \rho_{G(j)} \times \epsilon_{itj} \tag{3.7}$$

where ϵ_{itj} are GEV-distributed idiosyncratic preference for employer j, which may be correlated with preferences for similar employers in nest G but independent across time. Under the nesting structure, the probability of her choosing employer j can be decomposed as:[20]

$$p_{j|C} = \underbrace{p_{j|G(j)}}_{\text{choose } j \in G(j)} \times \underbrace{p_{G(j)|C}}_{\text{choose nest } G(j) \in C} \tag{3.8}$$

each of which is a function of (wage) offers within a choice set C. The choice probabilities enter the labor supply of incumbent employees and new workers from other employers, which influences the contracts set by employers.

At $t = 1$, workers enter the market for the first time and solve the problem in (3.7). At $t > 1$, a random fraction of incumbent workers can choose a new employer. One may consider an extension where workers can choose whether they search for new jobs or not. The probability of searching for new jobs, specified in (3.4), can be viewed as a reduced-form representation of workers' decision to re-enter the market, which may depend on market beliefs. But it would not substantially change the game between employers.

3.2.2 Employers' Problem

I focus on how employers set wages and allocate workers to innovation tasks in an intermediary period, $t \in \{2, , ..., T - 1\}$. The complete backward induction from $t = T$ to $t = 1$ can be found in Appendix A1.

At a period $t \in \{2, , ..., T-1\}$, employer j's value function is summed over incumbent employees

---

[20]I remove the subscript t since the problem of workers on the market is the same each period.

and potential recruits from other firms as follows:

$$V_{tj}\left(\bigcup_{\text{worker } i} I_{itj}\right) = \underbrace{\sum_{i:\, j(i,t-1)=j} v_{tj}^{(1)}(I_{itj})}_{\text{Incumbent}} + \underbrace{\sum_{i:\, j(i,t-1)\neq j} v_{tj}^{(0)}(I_{itj})}_{\text{Workers Outside}} \qquad (3.9)$$

where I_{itj} represents the information employer j has about worker i at the beginning of t.

The expected value from an incumbent employee is not the same as that from a potential new employee with the same public information for two reasons. First, information is asymmetric when an outside worker has a public innovation $y_{i(t-1)} = 1$ but whether it is high-quality $\widetilde{y}_{i(t-1)} \in \{0, 1\}$ is not publicly known yet. Second, the expected labor supply of an incumbent employee, denoted by $p_j^{(1)}$, takes a different form from the labor supply of a new worker, $p_j^{(0)}$.[21]

Contracts for Incumbent Workers

For an incumbent employee, employer j solves:

$$v_{tj}^{(1)}(I_{itj}) = max_{w,\tau}\, \underbrace{p_j^{(1)}(w; I_{itj})}_{\text{labor supply}} \times \underbrace{\left(MP_j(\widetilde{\pi}, \tau) + \beta \times E[v_{(t+1)j}^{(1)}(I)\,|\,I_{itj}, \tau] - w\right)}_{\text{MRPL at t and discounted continuation value, net wage}} \qquad (3.10)$$

where j's private belief $\widetilde{\pi} = Pr\left(H\,|\,I_{itj}\right)$ enters the expected marginal revenue product of a worker (3.1). Employers take into account the value from workers who will stay the next period, discounted by a common exponential factor β. The continuation value equals the value from an incumbent worker $v_{(t+1)j}^{(1)}$, expected over information next year (see 10.14).

Employers cannot observe who is on the market before setting the contract. Public belief $\pi = Pr\left(H\,|\,I_{it}\right)$ affects the chance of a worker searching for new jobs as well as offers from other employers. The expected labor supply from an incumbent employee, conditional on public information and taking as given the wages set by other employers, can be written as:

$$p_j^{(1)}(w, I_{itj}) = \underbrace{1 - \lambda_{G(j)}(\pi)}_{\text{off market}} + \underbrace{\lambda_{G(j)}(\pi) \times E_C[p_{j|C}(w, w_{(-j)})]}_{\text{on market \& choose } j \text{ again}} \qquad (3.11)$$

In comparison with monopsonistic wages in a static framework (e.g., Card et al. 2018), the dynamic wages are front-loaded with the expected continuation value from a job stayer, marked

---

[21]$p_j^{(1)} = p_j^{(0)}$ iff the information is symmetric (outside workers do not have a publication at $(t-1)$), and incumbent employees search for new jobs with probability $\lambda_{G(j)} \equiv 1$.

down by the inverse of labor supply elasticity $\xi_j^{(1)}$ (see 10.5).

$$w_{itj}^{(1)} = \underbrace{\left(MP_j(\widetilde{\pi}, \tau_{itj}^{(1)}) + \beta \times E[v_{(t+1)j}^{(1)}(I)|\, I_{itj}, \tau_{itj}^{(1)}]\right)}_{\text{MRPL plus continuation value}} \times \underbrace{\xi_j^{(1)} \times \left(1 + \xi_j^{(1)}\right)^{-1}}_{\text{markdown}} \tag{3.12}$$

Conditional on private belief $\widetilde{\pi}$ about an incumbent worker, employer j allocates her to innovation tasks to maximize the expected returns to innovation today, plus the continuation value. That is, employers take into account how task allocations would affect public information about a worker and her turnover tomorrow. Reducing the time she spends on innovation can lower the chance of her producing a publication and keep future wage bids from outside firms lower.

$$\tau_{itj}^{(1)} = max\{0, min\{1, \frac{1}{\zeta}(\ \underbrace{g_j \times q(\widetilde{\pi}) - 1}_{\text{returns to innov today}}\ + \beta \times \underbrace{\partial E[v_{(t+1)j}^{(1)}(I)|I_{itj}, \tau]/\partial \tau}_{\text{change in continuation value}}\ )\}\} \tag{3.13}$$

where the derivative of continuation value over task allocation, denoted by $\partial E[v_{(t+1)j}^{(1)}|...\tau]/\partial \tau$, may be negative if workers who produce an innovation are likely to be poached away (see 10.15 for details).

Contracts for New Workers

For an outside worker i from $j(i, t-1) \neq j$, employer j has access to information I_{itj} which does not indicate if a paper produced during $(t-1)$ is high-quality. The value function is therefore expected over the unknown quality $\widetilde{y}$: [22]

$$v_{tj}^{(0)}(I_{itj}) = max_{w,\tau} E_{\widetilde{y}}\,[\ \underbrace{p_j^{(0)}(w; I_{itj}, \widetilde{y})}_{\text{labor supply}} \times \underbrace{\left(MP_j(\widetilde{\pi}, \tau) + \beta \times E[v_{(t+1)j}^{(1)}(I)\,|I_{itj}, \widetilde{y}, \tau] - w\right)}_{\text{MRPL \& discounted continuation value, net wage}}] \tag{3.14}$$

where the labor supply expected from a new worker from nest G is:

$$p_{tj}^{(0)}(w; I_{itj}, \widetilde{y}) = \lambda_G(\pi) \times E_C\left[p_{j|C}(w, w_{(-j)})|\,\widetilde{y}\right] \tag{3.15}$$

When $\widetilde{y} = 1$, a worker is less likely to move from her previous employer to j. As shown in (10.6), an incumbent employer knowing $\widetilde{y} = 1$ revises upward the expected marginal revenue product and sets a higher wage, therefore reducing turnover.

Wages for new workers are marked down by the inverse of labor supply elasticity $\epsilon_j^{(0)}(\widetilde{y})$ (10.10), which like above is specific to the not-yet-revealed $\widetilde{y}$. In comparison with (10.5), the labor supply of new workers is more elastic with respect to wages than equally productive incumbent employees who may not search for new jobs at all. Therefore new workers receive a higher front-loaded wage

---

[22]The outer expectation over $\widetilde{y}$ can be removed if $y_{i(t-1)} = 0$, which implies $\widetilde{y} = 0$ and information is symmetric between employers.

than their incumbent counterparts. The wage gap may reflect a signing bonus in practice.

$$w_{itj}^{(0)} = E_{\widetilde{y}}[\underbrace{(MP_j(\widetilde{\pi}, \tau_{itj}^{(0)}) + \beta \times E[v_{(t+1)j}^{(1)}(I)|I_{itj}, \widetilde{y}, \tau_{itj}^{(0)}])}_{\text{MRPL plus continuation value} \mid \widetilde{y}} \times \xi_j^{(0)}(\widetilde{y}) \mid \text{enter j}] \times \left(1 + E_{\widetilde{y}}[\xi_j^{(0)}(\widetilde{y}) \mid \text{enter j}]\right)^{-1}$$

(3.16)

The optimal allocation of new workers to innovation tasks takes a similar form as that for incumbent employees (3.13):

$$\tau_{itj}^{(0)} = max\{0, min\{1, \frac{1}{\zeta}(E_{\widetilde{y}}[g_j \times q(\widetilde{\pi}) - 1 + \beta \times \underbrace{\partial E[v_{(t+1)j}^{(1)}(I)|I_{itj}, \widetilde{y}, \tau]/\partial \tau}_{\text{change in continuation value}} \mid \text{enter j}])\}\}$$

(3.17)

The difference from (3.13) is that employers have to take an expectation over the unknown $\widetilde{y}$. They also account for the adverse selection of movers in the sense that workers with $\widetilde{y} = 1$ are less likely to exit from their incumbent employers, due to a higher wage set by incumbent (e.g., Gibbons and Katz 1991; Greenwald 1986).

Appendix A1 completes the backward induction. At $t = 1$, all workers are new to employers, and information is symmetric. Employers therefore set wages and allocate workers according to a common prior about their ability upon labor market entry.

3.2.3 Equilibrium

In an imperfectly competitive labor market ($\frac{b}{\rho} < \infty$), firms set profit-maximizing wages conditional on information they have about workers and taking as given the wages set by other firms.[23] The subgame perfect Nash equilibrium with wages as strategic variables is defined as follows.

Definition 1 (Subgame Perfect Nash Equilibrium in Wages) *In the T-period game of wage posting in an imperfectly competitive labor market, the equilibrium comprises:*

- *$t = 1$ wages $\{w_{1j}(I)\}$ given each possible public information I about workers;*

- *$t > 1$ wages $\{w_{tj}^{(1)}(I_j)\}$ for incumbent employees, $\{w_{tj}^{(0)}(I_j)\}$ for workers from other firms, given each employer-specific information I_j about workers as assumed in (3.5);*

that are set by each firm to maximize its own profits, taking as given the wages set by other firms. Specifically, employers solve (10.16) at $t = 1$, (3.10,3.14) at $t = 2, ..., (T − 1)$, and (10.3, 10.8) at $t = T$. The expected labor supply is determined by workers who solve (3.7), conditional on the wages set by all potential employers.

---

[23]The assumption of simultaneous wage posting at the beginning of each period rules out ex-post bargaining between workers and firms, or offer matching. When setting wages, firms do not know which workers are on the market searching for new jobs at $t > 1$. Therefore they cannot "price discriminate" against those not looking for jobs.

As in the monopsony literature, the equilibrium wages set by firms in Bertrand competition are marked down from the net present values of a firm-worker match, by the inverse of labor supply elasticity:

$$
w^*_{itj} = \begin{cases} w^{(1)}_{tj}(I_{it} \cup \{\widetilde{y}_{i(t-1)}\}) & j = j(i, t-1) \text{, as in equations (3.12, 10.6)} \\ w^{(0)}_{tj}(I_{it}) & j \neq j(i, t-1) \text{, as in equations (3.16, 10.11)} \end{cases}
$$

The equilibrium task allocations set by employers are as follows:

$$
\tau^*_{itj} = \begin{cases} \tau^{(1)}_{tj}(I_{it} \cup \{\widetilde{y}_{i(t-1)}\}) & j = j(i, t-1) \text{, as in equations (3.13, 10.7)} \\ \tau^{(0)}_{tj}(I_{it}) & j \neq j(i, t-1) \text{, as in equations (3.17, 10.12)} \end{cases}
$$

Less innovative (lower-g_j) firms may not allocate any innovation tasks. But they can survive in this market if they are productive in routine activity ($f_j > 0$), and they will offer positive wages to workers. Assuming that the equilibrium wages are positive, the allocation of workers among firms can be shown to be unique at each period, given any possible information set.

Proposition 1 (Uniqueness under Monopsonistic Competition) *Given $b \in (0, \infty)$, $\rho_G \in (0, 1)$, $\lambda_G > 0$, and that $\forall j : f_j \in (0, \infty)$, equilibrium wages in Definition 1 are unique up to a nonzero scaling factor, and they result in unique allocation of workers between firms, conditional on firms' information about each worker. The probability of a worker choosing j at t in equilibrium:*

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

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

The proof in Appendix A2 closely follows Berry et al. (1995).

Thus far the labor market has been assumed to be imperfectly competitive $\frac{b}{\rho} < \infty$. Suppose labor supply is perfectly elastic at each period ($\frac{b}{\rho} \rightarrow \infty$ and $\lambda \equiv 1$), and information is incomplete but symmetric among employers. In this special case, the decision to allocate workers to innovation tasks is equivalent to the decision to provide general skill training that is transferable between firms.

Proposition 2 (Equilibrium under Public Information & Perfect Competition) *If the labor market is perfectly competitive ($\frac{b}{\rho} \rightarrow \infty$, $\lambda \equiv 1$) and information is always symmetric, each firm j offers a worker with*

public belief π:

$$\forall t : w_{tj}(\pi) = MP_j(\pi, \tau_{tj}(\pi)) \tag{3.18}$$

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

Proof: See Appendix A2.

There is no dynamic rent for any employer as workers are paid their full marginal revenue product of labor every period. The cost of innovation tasks is fully deducted from wages. Workers are not credit-constrained as they can always receive a positive wage by working on routine tasks (given $min\{f_j\} > 0$). The choice of $\tau_{tj}(\pi)$ is the first-best action in static equilibrium. Public learning, fully paid by workers, is equivalent to general training paid by workers in Becker (1964).

If the labor market remains perfectly competitive but information can be asymmetric as in (3.5), outside firms face a similar problem as in Henricks and Porter (1988). When there is no paper, information is symmetric and the equilibrium wages are as shown in Proposition 2. Conditional on observing a paper, however, outside firms is uncertain about the quality and adopt a mixed strategy to randomize their wage bids (Boozer 1994; Henricks and Porter 1988; Li 2012). Otherwise, there is always adverse selection (Greenwald 1986). It is unclear, however, if incumbent employers would allocate workers to innovation tasks efficiently.

To summarize, a firm's decision to learn a worker's research ability is in many ways similar to the decision to provide training. Firms are more willing to learn when information about workers is less public (Acemoglu and Pischke 1998), and when they have more monopsony power (Manning 2003; Stevens 1994) so that the learned talents are less likely to be poached. The benchmark model allows for both information and labor market frictions.[24] Next, I discuss the predictions from the subgame perfect Nash equilibrium in Definition 1, focusing on the implications of learning on job mobility that is observed in the data. Estimates

3.3 Model Predictions

The equilibrium under imperfect labor market competition (Definition 1) generates predictions on between-firm job mobility and innovation productivity as new information arrives, under the following key assumptions:

Assumption 1 (Labor Supply) *Labor supply is not perfectly elastic:* $b \in (0, \infty)$ *and* $\forall G : \rho_G \in (0, 1]$.

---

[24]Section 6 estimates the benchmark model in a balanced panel that covers the job transitions and innovation outputs by computer scientists in the first decade post PhD. Estimates in Table 9 suggest that the labor market is characterized by inelastic labor supply $\hat{b}/\hat{\rho} \ll \infty$, and $\lambda < 1$.

Assumption 2 (Dynamics) *At $t > 1$, there is a positive probability that each worker gets on the market and searches for new jobs: $\forall G \, \forall \pi : \lambda_G(\pi) > 0$. And the probability is non-decreasing in public belief π that a worker is H-ability: $\frac{\partial \lambda_G(\pi)}{\partial \pi} \geq 0$.*

Assumption 3 (Productivity) *All firms have positive productivity in routine tasks, and non-negative productivity in innovation: $\forall j : f_j > 0, g_j \geq 0$. H-ability workers are more productive in innovation tasks, $h > l$, and are more likely to produce high-quality innovation, $\widetilde{h} > \widetilde{l}$.*

The labor market is imperfectly competitive under Assumption 1 so that firms have wage-setting power. The positive probability that workers search for new jobs at $t > 1$ in Assumption 2 and firms' positive routine productivity under Assumption 3 yield positive wages (interior solutions) for any worker in equilibrium.[25] I state the three predictions below. See Appendix A3 for proofs.

Prediction 1 (Job Mobility in Response to Public Information) *Under Assumptions 1-3, conditional on public information about their ability, workers with any public innovation ($y = 1$) are*

a) *more likely to move to a new employer,*

b) *more likely to move to an employer with higher innovation productivity ($g_j \uparrow$)*

than coworkers without innovation ($y = 0$).

New innovation signals would improve the market belief that a worker is H-ability, under Assumption 3. Workers with higher market beliefs are more likely to re-enter the market and search for new jobs. Moreover, employers set higher wages for higher-π workers. As a result, workers with a public innovation are predicted to be more mobile between firms in (a).

Firms and workers are complementary in innovation. Firms with higher productivity g_j allocate more innovation tasks than other firms on average. It generates positive assortative matching (PAM) between high-π workers and high-g_j firms.

Prediction 2 (Job Mobility under Asymmetric Information) *Workers who have produced a high-quality innovation, i.e. $(y, \widetilde{y}) = (1, 1)$, are*

a) *less likely to leave an incumbent employer when the high quality $\widetilde{y} = 1$ is private information;*

b) *more likely to move and move upward after $\widetilde{y} = 1$ is revealed.*

than coworkers with outputs $(y, \widetilde{y}) = (1, 0)$.

---

[25]In Section 6, I relax these assumptions and estimate the model parameters instead. Estimates in Table 9 show that H-ability is more productive than L with $\hat{h} > \hat{l}$, $\widehat{\widetilde{h}} > \widehat{\widetilde{l}}$, consistent with Assumption 3. Higher-π workers from industry or postdoc markets are more likely to search for new jobs, in support of Assumption 2, but the reverse is true for tenure-track employees in academia.

The second prediction relies on the assumption that the quality of innovation, $\widetilde{y}$, is initially private information. Incumbent employer can set a higher wage based on a more favorable private belief $\widetilde{\pi} > \pi$ when $\widetilde{y} = 1$, and therefore reduces the turnover of $(1,1)$ workers relative to $(1,0)$ coworkers. Once $\widetilde{y}$ is revealed by the next period, however, Prediction 1 applies. $(1,1)$ workers are more likely to move and move up, once the market can recognize them as more productive than $(1,0)$ counterparts.

Prediction 3 (Productivity under Asymmetric Information) *Under Assumption 3,*

a) *Conditional on public belief, workers with initial outputs $(1,1)$ are more likely to produce innovation in the future than $(1,0)$ coworkers.*

b) *The gap between $(1,1)$ stayers and $(1,0)$ stayers is larger than that between $(1,1)$ movers and $(1,0)$ movers.*

On average, $(1,1)$ workers are more likely to be H-ability than $(1,0)$ workers. Therefore they are predicted to be more productive in future periods. Further, given the information structure (3.5), $(1,1)$ workers are initially known by the incumbent to be more productive than $(1,0)$ coworkers. $(1,1)$ stayers will be allowed to devote more time to innovation tasks than $(1,0)$ stayers. In contrast, $(1,1)$ movers and $(1,0)$ movers with the same public belief will spend an equal amount of time on innovation at a new employer. The difference in task allocation further magnifies the ability gap between $(1,1)$ and $(1,0)$, leading to (b).

The impact of asymmetric learning on total innovation productivity, however, remains ambiguous. Incumbent employers may allocate labor more efficiently given superior information. But $(1,1)$ workers may have moved up to more productive firms faster. I resolve the ambiguity by estimating the model and considering a counterfactual with symmetric information in Section 6.

In summary, the subgame perfect Nash equilibrium in this dynamic model of employer learning generates predictions on job mobility and productivity upon information revelation. I test these model predictions in the CS labor market in Section 5.

4   Data

I collected data on the career trajectories and research outputs of Ph.D. computer scientists. This section discusses the data sources, the matching between Ph.D. dissertation records and public

LinkedIn profiles, and the identification of on-the-job research that includes conference papers and matching patent applications.

4.1 Ph.D. Graduates in Computer Science

I focus on Ph.D. graduates in CS or closely related fields, who are qualified not only for tenure-track academic jobs but for research-intensive roles in the industry that require an advanced degree.[26] On the ProQuest Theses and Dissertation Database, I found about 81,000 Ph.D. dissertations in Computer Science or Electrical Engineering from the top 60 CS schools in the United States, between 1980 and 2021.[27] Each dissertation record provides the full name of the doctoral recipient, school and year of completion. Appendix Table B1 displays the number of dissertations by year. For school×year cells with particularly low or missing data on ProQuest, I collected about 15,000 more Ph.D. profiles from school-specific sources, such as department websites or dissertation repositories. For example, the number of new dissertations from Carnegie Mellon University dropped from 100 to 30 in 2014. I then collected additional dissertations from its own open-access repository KiltHub.[28] Figure 4 shows the total number of Ph.D. graduates in the sample by year, which stays around 3,000-3,300 per year from the top 60 schools since 2006.[29]

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

To gather information on the career progression of CS Ph.D.'s, I use public profiles on LinkedIn, the largest online professional network. My scraper program navigates the LinkedIn platform as a recruiter. For each person, I submit a web query that restricts to candidates from the same university, and then searches by the person's full name (with or without middle name), keywords that indicate a Ph.D. degree and keywords about computer science (see Appendix Figure B4). Among the profiles returned, the scraper checks if the full names are reasonably matched and if so, collects all public information such as profile summary, job history, education, and accomplishments that may include

---

[26]See Appendix Figure B1 for research scientist job ads. CS Ph.D.s may also work as engineers, but they often start as senior software engineers directly or as research engineers who also publish papers.

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

[28]See a detailed breakdown of dissertations found on ProQuest versus school-specific sources in Appendix Table B2. I did not look for other top 60 schools not shown in the table.

[29]NSF SER shows the average number of CS/EE Ph.D.s between 2010 and 2021 is 3,908. The gap includes Ph.D.s recipients out of the top 60 CS departments.

their publications or patents, and saves the profile picture if available.

A LinkedIn profile is considered fully matched to the PhD graduate only if the first name, last name, and PhD institution are matched exactly, and the year of Ph.D. completion is the same whenever it is available on the profile. About 51% queries returned at least one LinkedIn profile, and there are about 41,000 fully matched profiles in total.[30]

The matching rate is higher for more recent cohorts, as shown in Figure 4. LinkedIn was first launched in 2003, and its members grew from 37 million in 2009 to 875 million in 2023.[31] Each profile is formatted as a résumé, including sections on employment history and education background. I focus on employment records after a person receiving her Ph.D. degree. On average there are 2.1 post-Ph.D. employers from the industry, 0.3 academic (tenure-track) employers and 0.2 postdoc employers in a matched profile (Table 2).

I construct a longitudinal dataset of post-Ph.D. employment history for over 40,000 fully matched LinkedIn profiles. For each person×year, I record the primary employer and job title.[32] The person×year panel has about 647,000 observations in total.[33] 94% of the observations have a nonmissing employer.[34] A job-to-job movement in year t is defined as a change in one's primary employer in comparison with her employer next year: $j(i,t) \neq j(i,t+1)$. Years without any employer would not be considered as a job movement. About 12% of workers at non-top firms move to a new employer per year, whereas about 7% of those at top firms or in academia move (Table 3).

Despite the popularity of LinkedIn as the largest professional network today, there might be important selection of Ph.D.s into the matched sample. Computer scientists who always work in academia are admittedly less likely to ever use LinkedIn. I aim to complement it with academic

---

[30]With the Recruiter Lite account, LinkedIn allows me to view public profiles within my third degree of connections. To deal with the limitation, I actively connected with a random sample of Ph.D. graduates before the web scraping for each school. I also connected with individuals who published at CS conferences, or worked as research scientists at various companies. If an individual is on LinkedIn but falls outside my 3rd-degree connections, the search result would return a message "Out of Network". There are roughly 1,800 out-of-network profiles in total, out of fifty thousand queries that return at least one profile on LinkedIn. Based on a random sample, it appears that most individuals out of network have less than 100 connections on LinkedIn. I am working on another round of data collection to see if I can collect their profiles.

[31]Business of Apps Analysis https://www.businessofapps.com/data/linkedin-statistics/

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

[33]For each person I include all years between the year of Ph.D. and the year on her most recent job. If the dates for a job says "2019-Present", I will keep all years up to 2022. 96% of matched individuals have a non-missing record (including retirement or unemployment) up to 2022.

[34]About one thousand person×year rows indicate self-employment as the primary employment and are not included in the analysis.

CVs of CS professors in future work. This paper focuses on career progression of workers who have some experience in the industry or move between industry and academia. This group is more likely to have a LinkedIn profile and update it when they move between employers or get promoted.

4.3   Research Production: Papers and Matched Patents

4.3.1   Research Papers

To measure the research productivity of Ph.D. computer scientists, I collected papers that are published as proceedings in 80 CS academic conferences and two machine learning journals, which are used to rank CS institutions across all areas in CSRankings. CS conferences have a relatively short history.[35] I focus on publications since 2000.

The main data source of research papers is Scopus, an abstract and citation databases of peer-reviewed literature produced by Elsevier.[36] Using Scopus Search API, a query is submitted for each conference/journal×year. It returns a list of papers with titles, author names and identifiers, as well as affiliations, which should include the employer of an author at the time of publication.

In addition, I collected papers from DBLP, a computer science bibliography website with a designated page for each conference×year or journal×year. DBLP provides a comprehensive list of papers across conference×year's. DBLP makes it easier to differentiate between conference proceedings from non-research publications such as lecture notes, which might have been included by Scopus. However, DBLP does not provide author affiliations as Scopus does.

To match papers with individuals' education and employment history, I developed a script to clean and harmonize the names of author affiliations from Scopus, and the names of Ph.D. schools and employers from LinkedIn. A paper matched to an author's Ph.D. institutiton by (author name, affiliation, year of publication) is labeled as pre-Ph.D. research. After Ph.D., a paper is considered as on-the-job research if the author affiliation matches with her incumbent employer at the time of publication. The publication cycle is significantly shorter in computer science. It is unlikely for a dissertation chapter to be published as a conference proceeding years later.[37] About 28% of matched

---

[35]For example, the International Conference on Learning Representations, one of the leading conferences in machine learning, was established in 2013.

[36]I am especially grateful to Anna Le Sun (Berkeley/Stanford) for her help with the large-scale data collection from Scopus.

[37]To differentiate between pre-Ph.D. works and post-Ph.D. on-the-job research, I check if coauthors on a paper are affiliated with the Ph.D. school or with the current employer. Roughly 1% of post-PhD publications have the majority of

computer scientists have at least one on-the-job research publication post Ph.D. (Table 2).

Table 3 summarizes the person-year level panel data that merges employment history from LinkedIn with the publication records above. To be consistent with the empirical analysis, I restrict to full-time employment records after 2000 and summarize the data by whether a worker is currently at a non-top firm, top firm or academia. Per year about 2.3% of workers at non-top firms, 10.3% at top firms, and 18.4% in academia have at least one CS paper.[38]

4.3.2   Matched Patent Applications

Firms often resort to traditional intellectual property protections for inventions that are disclosed in a research paper. I first establish a paper-patent linkage if the following conditions are satisfied:

1. The majority of authors on the paper are also inventors on the patent application, and vice versa.
2. A patent assignee can be matched to an author's affiliation on the paper, which is also her current employer as shown on LinkedIn.
3. The patent application is initially filed between $[-2, 1]$ years relative to the publication of the paper (using conference date).[39]
4. Content is similar: the distance between a paper's (title, abstract), transformed into a vector via GPT4-ada embedding model, and a patent application's is less or equal to 0.35.

For each paper, I sort potential patent matches by the number of shared team members, the text similarity in abstracts (as a proxy for similarity in inventions), and the time difference between the earliest filing of a patent application and the publication date of the paper, in ascending order. I use the first patent application returned as the best possible match.

About 24% of papers by matched computer scientists from industry, and 5% of papers by those from academia, are accompanied by a patent application. 90% of the matched patent applications are filed before the research paper shows up at a conference, and the other 10% are filed within 12 months. Table 1 provides a few paper-patent matches. I selected a random sample of matches, and the vast majority are almost perfect matches with similar research ideas and findings. Patent

---

coauthors affiliated with the Ph.D. school, and are excluded from on-the-job research production.

[38]Workers in academia include research staff or engineers who work for a university but may not be active in research as postdoc's or faculty.

[39]The patent laws in the U.S. allows the inventors to apply for a domestic patent for inventions that are disclosed in any publication no earlier than a year ago. In most other countries, inventions that have been disclosed, for example via a research paper, cannot be filed as a patent application.

applications, however, often contain more technical details and are more precise about contributions that can be claimed as inventions than what one can observe from a paper alone. I use the high-quality matches that satisfy all four conditions above for the main analysis, and show the results are robust if I add more matches by relaxing condition 4, which may double the match rates[40]

Papers with a matched patent are higher-quality ona v. In the first year, they receive roughly the same citations as those without a matched patent. But in 3-5 years, the citation difference expands to 25-60%, as shown in Table 4 and Figure 2. The quality difference between papers with and without a matched patent is important. If patenting decision is completely random, it is not clear if an incumbent employer can identify more influential works earlier than the outside market. The divergence in the citations suggests: 1) incumbent employers have additional information about the quality of a paper and can act on it by filing for a patent, 2) it takes time for the outside market to recognize valuable research and the timing of the divergence in citations is consistent with the revelation of a matched patent application.

5 Empirical Tests for Employer Learning

I test for the model predictions 1, 2 and 3 that detect employer learning from job mobility and productivity patterns in the panel data of computer scientists. There is evidence of public learning (Prediction 1) as job mobility increases for workers with a publication, relative to similar coworkers. To test for asymmetric learning, I leverage the delayed disclosure of whether a paper has been filed as a patent application. When information is initially asymmetric, workers with both a paper and a matched patent move at a slightly slower rate than those with a paper only. But mobility increases significantly when a lagged patent application becomes public information, consistent with Prediction 2. I discuss the implications of employer learning on productivity in the test for Prediction 3, which indicates that movers are adversely selected and outside firms allocate workers inefficiently under asymmetric information.

---

[40]If condition 4 is relaxed to 0.50 as a threshold, there can be almost 10,000 paper-patent links, doubling the match rates in both academia and industry. However, the matched patent application can be on a related but not the same subject as the paper. Related patents can be valuable signals as well, but it is unclear if they represent the asymmetric information between incumbent and outside employers conditional on the public information in a paper.

5.1 Baseline Specification

To test for employer learning in Predictions 1-2, I compare the job mobility of workers who have new public signals such as a paper or a matched patent application with that of similar coworkers without new signals at the same time. Equation 5.1 presents the baseline specification for testing the first two model predictions on mobility responses to new information:

$$M_{it} = \underbrace{\beta_{10}\, d_{it}(1,0) + \beta_{11}\, d_{it}(1,1)}_{\text{based on new } (y_{it},\widetilde{y}_{it})} + \underbrace{\gamma_{10}\, D_{it}(1,0) + \gamma_{11}\, D_{it}(1,1)}_{\text{based on lagged outputs}} \tag{5.1}$$

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

where M_{it} is a mobility outcome at person×year level, including any move between employers, any move into a top firm, or a promotion. Define $y_{it} = 1$ if worker i has any CS paper released in year t, and $\widetilde{y}_{it} = 1$ if she has a patent application matched with the new paper. There are three potential outputs: $(y_{it}, \widetilde{y}_{it}) \in \{(0,0),(1,0),(1,1)\}$. I define an indicator $d_{it}(p,q) = 1$ if $(y_{it}, \widetilde{y}_{it}) = (p,q)$. Since patent information $(\widetilde{y}_{it})$ becomes public with a delay, this specification allows outcomes to vary with lagged innovation outputs. Whether there has been a matched patent application for papers between $[t-3, t-1]$ is considered as public information at t. $D_{it}(p,q) = 1$ if her lagged outputs satisfy: $\mathbf{1}\left[\sum_{s=t-3}^{t-1} y_{is} > 0\right] = p$ and $\mathbf{1}\left[\sum_{s=t-3}^{t-1} \widetilde{y}_{is} > 0\right] = q$, for $(p,q) \in \{(0,0),(1,0),(1,1)\}$.

The reduced-form strategy compares workers with new productive signals with similar coworkers without a signal. The firm-year fixed effects, denoted by $\phi_{j(i,t),t}$, absorb firm-specific shocks such as a layoff, and allow us to compare workers within the same firm. W_{it} is a vector of worker characteristics such as educational background (bachelor and Ph.D.), gender (from first names or profile pictures), and time-varying controls such as a polynomial of experience since Ph.D. and position types (e.g., engineers vs. scientists).

The coefficients of interest are $(\beta_{10}, \beta_{11}, \gamma_{10}, \gamma_{11})$. β_{10} captures the difference in outcome M_{it} between $(1,0)$ workers with a new publication but no patent, and $(0,0)$ workers without a new paper as base. β_{11} captures the difference between $(1,1)$ workers and $(0,0)$. γ_{10} represents the gap between lagged $(1,0)$ workers who have at least one paper but no matched patent between $[t-3, t-1]$ and lagged $(0,0)$ workers without any paper during that time. And γ_{11} compares the lagged $(1,1)$ who

also have a matched patent with the lagged $(0,0)$. The model predictions can be written as follows:

Prediction 1 $\rightarrow \beta_{10} > 0$, $\beta_{11} > 0$ for mobility $M_{it} \in$ {Any Move, Upward Move}

Prediction 2 $\rightarrow \beta_{11} < \beta_{10}$, whereas $\gamma_{11} > \gamma_{10}$ and $\gamma_{11} > 0$, for mobility M_{it}

The following sections present the regression results, discuss alternative specifications, and adjust the baseline specification to test for Prediction 3 regarding the productivity differences between movers and stayers.

5.2 Mobility Responses to New CS Papers

I estimate 5.1 over a person-year level panel for matched CS Ph.D.s. The estimation sample is summarized in Table 3, and for each person it includes years of full-time employment post Ph.D. and post 2000.[41] Computer scientists at non-top firms are more likely to move between employers than those at top firms or in academia (Table 3 and Figure 3). Motivated by this fact, I estimate separate regressions by whether a person is currently employed by non-top Firms, Top Firms or Academia, throughout this section.

The first three columns of Table 5 reports the OLS estimates of 5.1 for the dependent variable $M_{it} = 1[j(i, t+1) \neq j(i, t)]$, defined as any move between firms.[42] $(1, 0)$ workers with a publication but no patent from non-top firms are 3.5 ppt($t \approx 6$) or 27% (Table C1) more likely to move than $(0, 0)$ workers without any output, and $(1, 1)$ workers are 1.8 ppt's or 13% more likely to move.[43] The mobility difference between $(1, 0)$ and $(0, 0)$ is much smaller for workers from top firms or academia, and the difference between $(1, 1)$ and $(0, 0)$ is insignificant from 0 for both. Taken together, there is evidence for Prediction 1(a) outside the top firms or academia that workers with new innovation are more mobile in the labor market.

Prediction 1(b) further states that workers with new public signals are more likely to move into

---

[41]There are employment records before 2000 for earlier Ph.D. cohorts but I did not collect publication records. As discussed in Section 4.3, CS conferences have a relatively short history.

[42]Estimates of other regression coefficients are shown in Table C0.

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

$$E[M_{it}|d_{it}, D_{it}, W_{it}] = exp\left(\beta_{00} + \beta_{10}\, d_{it}(1,0) + \beta_{11}\, d_{it}(1,1) + \gamma_{10}\, D_{it}(1,0) + \gamma_{11}\, D_{it}(1,1) + \text{controls} + \phi_{j(i,t),t}\right) \quad (5.2)$$

which includes the same set of controls and firm-year fixed effects as in linear 5.1. The estimates for β_{10}, β_{11} are interpreted as proportional increases in job mobility from the baseline $(0,0)$ group.

more productive firms. Top firms (Google, Facebook, IBM, Microsoft, Amazon, Apple) are arguably more prodyctive than an average employer outside. I consider moving into a top firm from other employers in the industry as an upward move.[44] Columns (4)-(6) of Table 5 show the results for outcome defined as being employed by a top firm next year. Workers with a new publication from non-top firms are significantly more likely to move to a top firm: the gap in such upward mobility between $(1, 0)$ workers and $(0, 0)$ workers is around 1.8 ppt's (51% difference in Table C1), and the gap between $(1, 1)$ and $(0, 0)$ is around 1.3 ppt's or 28%. New publications do not seem to have a significant effect on the probability of staying at top firms, for workers who are already employed by top firms.

It is consistent with the model prediction that both mobility responses above are stronger outside the top. Equilibrium wages are increasing in a firm's (innovation) productivity. Workers who are revealed to be good researchers are more easily lured away by more productive employers that can offer a higher wage.

5.3   Mobility Responses to Initially Private Patent Applications

The test for asymmetric employer learning leverages the delayed disclosure of a patent application matched with the paper. As discussed in Section 4.3, roughly 24% of CS papers with authors from the industry, and 5% of papers from academia are accompanied by a patent application. Papers with a matched patent are also better cited in the future (Table 4; Figure 2).

Prediction 2 says $(1, 1)$ workers are initially less likely to leave incumbent firms than similar $(1, 0)$ coworkers ($\beta_{11} < \beta_{10}$), but they are more likely to move and move up when information becomes symmetric ($\gamma_{11} > \gamma_{10}$ and $\gamma_{11} > 0$). Table 5 estimates that $(1, 1)$ workers at non-top firms are 1.7 ppt's less likely to move than $(1, 0)$ workers in the same year, but the difference is not statistically significant. $\hat{\beta}_{11} < \hat{\beta}_{10}$ also holds for workers from top firms or academia, but the difference remains insignificant.

There is stronger evidence for part (b) of Prediction 2 that workers with lagged outputs $(1, 1)$ are more likely to move and move up than workers with lagged $(1, 0)$. Whether a research paper from the previous three years has been filed as a patent application should be public information

---

[44]This definition of upward mobility is imperfect. There may be smaller, more innovative firms than the tech giants. I show results on alternative upward mobility outcomes in Table 6.

at t. As shown in the first column of Table 5, workers with lagged $D_{it}(1,1) = 1$ at non-top firms are 1.8 ppt's or 13.1% more likely to move between firms at t relative to workers without lagged publications. Workers with lagged papers but no matching patents $D_{it}(1,0) = 1$ in the industry are no more likely to move than the base group with $D_{it}(0,0) = 1$ group. The difference between $\widehat{\gamma}_{11}$ and $\widehat{\gamma}_{10}$ is negligible for workers from top firms or academia. It suggests lagged patent application, as a signal that comes in with some delay, is more important for workers at non-top firms. The findings are similar for the upward mobility outcome in columns (4)-(6) of Table 5 and Appendix Table C1: lagged $(1,1)$ workers are 1.3 ppt's or 35% more likely to move from non-top to top firms than lagged $(0,0)$, whereas lagged $(1,0)$ workers are 0.5 ppt's or 16% more likely to move up (relative to the base group without a paper during $[t-3, t-1]$).

In contrast, whether a recent paper has been matched to a patent application does not matter for workers in academia. Less than 5% of papers by academic researchers are accompanied by a patent, which may represent collaborations with the industry and matter less for tenure evaluation within academia. For workers moving from industry to academia, patents also do not matter as much as they do in the industry. Table 6 presents an OLS regression of employment by an academic employer next year on research signals as in 5.1. Workers with a new paper, across all three groups, are significantly more likely to move into / stay in academia. Having a recent paper in the past three years also increases the mobility from non-top firms to academia significantly, more so for the lagged (1,1) group but the difference from the lagged (1,0) is insignificant. Admittedly, moving from industry to academia does not happen very often. However, given rising concerns about the poaching of CS professors by tech companies (e.g., Gofman and Jin 2022), it is worth knowing if increasing academic jobs for industry workers can help counter the brain drain from academia and increase research productivity. I will study this policy-relevant question formally in Section 6.

5.3.1 Robustness Checks

As a robustness check, I replace indicators for lagged outputs D_{it} by an indicator for any patent application published by the USPTO in year t, which captures the revelation of a recent $(1,1)$ based on USPTO records. Table 7 shows that between-firm job mobility increases by 3.3 ppt, and upward mobility into top firms increases by 1.3 ppt, for workers with a new patent publication at non-top

firms, similar to the estimate $\widehat{\gamma}_{11}$ on $D_{it}(1,1)$ for lagged outputs in column (1).

One may be concerned that patent applications that are not matched to a CS paper can also matter for job mobility and confound the estimated mobility differences between $(1,0)$ and $(1,1)$ workers. To address this concern I add controls for whether a worker has a patent application in year t or recently during $[t-3, t-1]$. Regression estimates remain largely unaffected. Authors from nontop firms who recently produce $(1,1)$, for example, are 1.7 ppt's or 12% more likely to move between employers, and 1.2 ppt's or 32% more likely to move to a top firm the next year than $(1,0)$ authors when they can be differentiated in the labor market (Appendix Table C3). Workers who have other patent applications (not matched to papers) at non-top firms are 0.6 ppt more mobile than their coworkers. They are also more likely to move to or stay at top firms (columns 4-5 in Table C3). But these effects are smaller relative to that of CS papers or patents matched to CS papers. The differences in inter-firm mobility between $(1,0)$ and $(1,1)$ workers are unlikely to be driven by other patenting activity.[45]

One may also be concerned that $(1,1)$ workers who have a paper and a matched patent stay longer to wait for a patent grant. However, it often takes 3 or more years for a patent application to be finally approved and granted. There is limited evidence that employers share rents from a granted patent with inventors, unless they are managers at firms that just began to patent (e.g., Kline et al. 2019). Workers who have been acknowledged as inventors on a patent application also would not be removed from it during the examination process, regardless of whether they leave the firm or not. Table 7 further shows that a newly granted patent does not increase job mobility for any group. CS research especially in AI-related fields progresses much faster than the patenting process. It is not surprising signals about a person's latest research matter more for her job mobility.

The empirical evidence of Prediction 2 is also robust to model specification. Appendix Table C1 presents Poisson regressions of job mobility on innovation signals, as specified in 5.2. Lagged $(1,1)$ workers at non-top firms are 12% more likely to move, and 19% more likely to move into a top firm the next year relative to the lagged $(1,0)$ workers.

So far the comparison has been done between observably similar coworkers. I add person

---

[45]Patenting has remained largely stable over the past two decades. The share of patents that are matched to CS papers, however, has been increasing especially since 2015, which is not driven by an expansion of matched LinkedIn profiles (see Appendix Figure B6). I plan to study a firm's decision to patent vs. publish a paper, and the complementarity between these two practices in future work.

fixed effects into the linear model 5.1 to estimate within-person changes in job mobility in response to research signals. Among workers at non-top firms, having a new paper, regardless of whether it comes with a patent, increases her job mobility by 3.5 ppt's relative to a year without any paper (Appendix Table C2). Any lagged patent with a matched patent further increases job mobility in this group by 3.6 ppt, which is $\hat{\gamma}_{11} - \hat{\gamma}_{10} = 2.8$ ppt's higher than the lagged $(1,0)$ group who only have paper(s). The second set of results in Table C2 further highlight that within person $\hat{\beta}_{10} \approx \hat{\beta}_{11}$, different from the estimates between coworkers in Table 5. But lagged $(1,1)$ workers from non-top firms are significantly more likely to move and move upward into a top firm, supporting part (b) of Prediction 2 that $\hat{\gamma}_{11} > \hat{\gamma}_{10}$.

5.3.2 Other Mobility Outcomes

Another upward mobility outcome is defined by moving into a higher-wage firm. Using labor condition applications filed by employers for H1-B workers, I compute an average wage per employer×yr for engineers and research scientists and merge it into the panel. Workers with a new paper at non-top firms are 3.4 ppt's more likely to move into a higher wage firm (Table 6. Lagged research works with matched patents also increase mobility into higher-wage employers consistently across groups.[46]

New research signals, including both a paper this year or a patent application matched with recent papers, increase the promotion rates across all three groups. Promotions are defined based on job titles, and in academia, it includes getting tenure. Under asymmetric information, Waldman (1984) suggests the higher ability workers, the new $(1,1)$ group in this case, are less likely to get promoted so the outside market cannot identify them from the promotion.[47] There is conflicting evidence from non-top vs. top firms, however. $(1,1)$ workers are 1.3 ppt's more likely to get promoted than $(1,0)$ at non-top firms, but are 0.9 ppt's less so at top firms. Neither difference is significant. It remains difficult to tell if employers are also hiding talent by delaying promotions in this setting. I further break down the changes in job titles Appendix Table C4. Workers with a new paper, and

---

[46]Staying at an employer where the real wages are growing would also be coded as 1 for the dependent variable $M_{it} = \overline{w}_{j(i,t+1),t+1} > \overline{w}_{j(i,t),t}$

[47]Hagele (2022) finds evidence of talent hoarding by managers within a firm. It is possible that some workers do not get an opportunity to publish at all under managers who hoard talent. But it is beyond the scope of this paper. I focus on the margin of asymmetric information that conditional on publishing, $(1,1)$ cannot be differentiated from $(1,0)$ by the outside market initially.

lagged $(1, 1)$ workers with a newly revealed patent application, are more likely to become a research scientist the next year. The effects are smaller or negative on becoming an engineer or manager, in comparison. Research scientist is a higher-paid job than an engineer and may be considered as a form of promotion.

To summarize, among workers at non-top firms there is strong evidence of delayed job mobility under asymmetric information (Prediction 2). Before the revelation of a patent application matched to the paper, $(1, 1)$ workers are less mobile than $(1, 0)$ workers: $\widehat{\beta}_{11} < \widehat{\beta}_{10}$ (Table 5) supports Prediction 2(a) but the difference may be statistically insignificant. Workers with lagged $(1, 1)$, once distinguishable from workers with lagged $(1, 0)$, are more likely to move and move into top firms or academia, or get promoted, in support of 2(b).

5.4 Productivity under Asymmetric Information

The benchmark model suggests employers have a stronger incentive to learn when some of the information can be kept private (Proposition 1). Incumbent employers with additional information can also allocate workers between routine and innovation tasks more efficiently. Quantifying the aggregate impacts of asymmetric learning on productivity requires estimating the full structural model with endogenous sorting and task allocations, which are presented in the next section. Here I focus on testing Prediction 3 that $(1, 1)$ workers are on average more productive than $(1, 0)$ workers, and the difference among stayers is larger than that among movers. I estimate a Poisson regression as follows:

$$E[Y_{it}|d_{it}, W_{it}, j(i,t)] = exp\big(\underbrace{(\delta_{00,S} + \delta_{10,S} \times d_{it}(10) + \delta_{11,S} \times d_{it}(11))}_{\text{Stayers}} \times (1 - M_{it}) \tag{5.3}$$

$$+ \underbrace{(\delta_{00,M} + \delta_{10,M} \times d_{it}(10) + \delta_{11,M} \times d_{it}(11))}_{\text{Movers}} \times M_{it} + W_{it}' \Gamma + \phi_{j(i,t),t} \big)$$

where Y_{it} is a discrete count of papers or patents in the next three years as a proxy for future productivity. The base group comprises $(0, 0)$ workers who do not have a new paper and stay at the each's incumbent employer ($M_{it} = 0$). The vector of coefficients $(\delta_{10,S}, \delta_{11,S})$ represents the proportional increase in productivity of the $(1, 0)$ and $(1, 1)$ stayers from the base group, respectively.

Similarly, $(\delta_{10,M}, \delta_{11,M})$ captures the proportional increase relative to $(0,0)$ movers. Additional controls W_{it} and firm-year fixed effects $\phi_{j(i,t),t}$ are the same as specified in 5.1. Prediction 3 can be translated into:

$$a) \; \delta_{11,S} > \delta_{10,S}, \text{ and } \delta_{11,M} > \delta_{10,M}$$
$$b) \; \delta_{11,S} - \delta_{10,S} > \delta_{11,M} - \delta_{10,M}$$

Table 8 shows that $(1,1)$ workers on average produce more papers and papers with a matched patent than $(1,0)$ coworkers, regardless of staying at the incumbent employer or not. The differences among stayers are significantly positive in all three groups of incumbent employers. For example, $(1,1)$ stayers at non-top firms produce 25% more papers and 62% more papers with a matched patent than $(1,0)$ stayers, conditional on observables and firm-year fixed effects. The estimates provide evidence for Prediction 3(a). Workers who produce $(1,1)$ are more likely to have high research ability than workers who produce $(1,0)$, resulting in the productivity differences observed among both stayers and movers. There is more noise in the estimated difference among movers (second column of Table 8), indicating a large heterogeneity between destination firms that may

Part (b) of Prediction 3 says the gap between $(1,1)$ and $(1,0)$ stayers is larger than that between movers. I find evidence of this prediction by comparing the two columns of Table 8. $(1,1)$ movers produce 13% more papers, or 25% more papers with a patent than $(1,0)$ movers, both of which are smaller the gap among stayers shown in the first column. Admittedly, the estimates for movers are noisier, suggesting a large heterogeneity in destination firms that may attenuate the productivity difference between the two groups.[48] The diff-in-diff in the production of papers with a matched patent, however, is larger and marginally significant for workers from non-top firms or academic employers.

The contrast between movers and stayers can be explained by incumbent employers allocating workers more efficiently given superior information. When an incumbent employer can tell $(1,1)$ stayers apart from $(1,0)$, it can allow $(1,1)$ to devote more time to research tasks in the future, further expanding the productivity difference from $(1,0)$'s. In contrast, the outside market views $(1,1)$ and

---

[48]Under asymmetric information, $(1,0)$ are more mobile than $(1,1)$ (see Table 5). A $(1,0)$ mover who enters a more productive firm may be more productive than $(1,1)$ mover who enters a less productive firm.

$(1, 0)$ workers as the same, and cannot fully exploit their ability difference as the incumbent does (see Appendix A3 for proof).

The results are also related to adverse selection under asymmetric information. As shown in Table C5, movers are estimated to be less productive than movers on average, for all groups except for some movers out of academia.[49] Adverse selection of movers is widely documented in the learning literature (e.g., Gibbons and Katz 1991; Henricks and Porter 1987). However, adverse selection alone cannot explain part (b) of the prediction that the productivity gap is larger for stayers than for movers.

In summary, this section presents empirical evidence of employer learning, especially asymmetric learning when an incumbent employer has information about matched patents earlier than the outside labor market. Mobility between firms is delayed for workers who produce a paper with a matched patent that is not revealed initially. Once information about patents is revealed, $(1, 1)$ workers move more frequently between firms and are more likely to move from a non-top firm to a top firm than $(1, 0)$ coworkers without a matched patent application. As for the impacts of asymmetric learning on productivity, incumbent employers are estimated to allocate workers more efficiently given superior information. However, higher-ability workers who produce hidden $(1, 1)$ might be able to move to productive firms earlier if the information were symmetric between employers. The aggregate impact on productivity remains ambiguous from the empirical tests. I tackle this challenge by estimating the structural model in the next section.

6 Quantitative Analysis

Would reducing information asymmetry between employers increase aggregate talent revelation and innovation outputs? This question is not fully answered by the empirical evidence of asymmetric learning so far. Employers may be reluctant to allocate workers to innovation tasks if all productive signals become public information. On the other hand, higher-ability workers could move to more productive employers faster under more symmetric revelation. These forces generate ambiguous predictions on productivity and longer-term talent revelation. To answer this question, I

---

[49]Column (6) in Table C5 shows $(1, 0)$ movers out of academia produce more patents in the next three years. It does not indicate this group is positively selected. As discussed in Section 4, patenting is more common in the industry than in academia. It is not surprising movers from academia to industry produce more patents.

estimate the benchmark model in Section 3, which features asymmetric learning between heterogeneous employers in a monopsonistic labor market. Given the estimates, I consider a counterfactual scenario where matched patent applications are disclosed simultaneously as research papers. I find an increase in job mobility among productive workers, a 6% increase in the revelation of high-ability workers, and a 5% increase in total innovation. The changes in the equilibrium outcomes provide estimates of the impacts of asymmetric learning (at the margin of patenting).

6.1 Empirical Strategy

Following the employer learning literature, I assume most of the learning takes place in the first decade of a worker's career (e.g., Altonji and Pierret 2001, Farber and Gibbons 1996). I estimate the model on a balanced panel of workers who graduated with a CS/EE PhD between 2000 and 2012 and have nonmissing full-time employment records in the first ten years post-PhD (see Table 2 for a sample overview). The goal is to find model parameters that maximize the joint likelihood of job movements between employers and innovation outputs (papers only, or papers with a matched patent) throughout the first decade.

6.1.1 Assumptions and Likelihood Function

Timing. The model has ten periods that correspond to 1-10 years post PhD (model timeline in Section 3.1.4).

Ability. Given information about a worker's research records across the first 10 years since Ph.D., I consider a worker to be high-ability, $\hat{\alpha}_i = H$, if her cumulative citations place her in the top 10% among the Ph.D.'s in sample.[50] Throughout the first decade, however, employers are assumed to be uncertain about the workers' true ability and update their beliefs based on observed signals.[51]

Priors. Employers form a common prior based on initial information at the beginning of $t = 1$. I assume the prior is drawn from a mixture of two normal distributions:

$$ln(\frac{\pi_{i1}}{1 - \pi_{i1}}) \overset{d}{\sim} \hat{z}_{i1} \times \mathcal{N}(\overline{\mu}, \overline{\sigma}) + (1 - \hat{z}_{i1}) \times \mathcal{N}(\underline{\mu}, \underline{\sigma}) \tag{6.1}$$

---

[50] $\hat{\alpha} = H$ if a worker's cumulative 3-yr citations exceed 50, or cumulative 5-yr citations exceed 100, or cumulative citations (no time window) exceed 150. Based on these thresholds, 1,500 Ph.D.'s out of 12,829 in the balanced sample are labelled as high-ability.

[51] Employers' belief would converge to 1 if a worker is revealed to be H, or 0 if she is revealed to be L.

where mixture weights are calibrated as $\hat{z}_{i1} = Pr(\hat{\alpha} = H \mid I_{i1})$. Initial information I_{i1} includes Ph.D. school, initial nest of employer $G(j(i, 1))$ and whether the first job is a research position. For example, for a graduate of the top 25 CS programs and working on the tenure track immediately, her initial prior is drawn from $\mathcal{N}(\overline{\mu}, \overline{\sigma})$ with probability 0.33, which is the fraction of $\hat{\alpha} = H$ in this initial group of workers (see Table D1). I refer to $\theta = (\overline{\mu}, \overline{\sigma}, \underline{\mu}, \underline{\sigma})$ as the nuisance parameters that govern the distribution(s) of initial prior, but do not affect the equilibrium labor supply, wages and task allocations in the model.

Employers. Denote by j_{it} the primary employer (group) of worker i in period t. There are over seven thousand unique employers in the balanced panel of workers. I group the employers as follows and allow heterogeneous productivity between groups but not within[52]:

    I Academia - Tenure-track
          1. Tenure-Track at non-top Schools
          2. Tenure-Track at Top CS[53]

   II Academia - Postdoc
          3. Postdoc at non-top Schools
          4. Postdoc at Top CS[54].

  III Industry - Top Firms
          5 - 10: IBM, Microsoft, Amazon, Facebook, Apple, Google

  IV Industry - Non-top Firms
          11 - 16: By Location - San Francisco Bay Area, West Coast (other than SF), East Coast, Other Locations in the U.S., Foreign Locations, Missing.[55]

Employers within each group j are assumed to have the same productivity (f_j, g_j). The 16 groups of employers (henceforth "employers") belong to four higher-level nests, denoted by $G(j) \in$ {Tenure-track, Postdoc, Top Firms, non-top Firms}.

---

[52]This grouping is equivalent to assuming that employers within a group are perfect substitutes to workers, i.e. diversity between employers within a group is not valued, as remarked in Dixit and Stiglitz (1977). Under this assumption, workers would draw an idiosyncratic preference for the group as a whole, and the share of workers in a group is independent of the number of firms within it. I am considering alternative grouping in ongoing analysis.

[53]Top CS includes the top 25 CS departments ranked by CSRankings: CMU, Berkeley, Stanford, MIT, Georgia Tech, Cornell, USC, UIUC, Princeton, Washington State, UCLA, UCSD, UMass - Amherst, UMich, Purdue, Maryland, Northeastern, Madison, Columbia, UT-Austin, UPenn, NYU, UC-Irvine, UC-Santa Barbara, UChicago, Stony Brook.

[54]A university can be both a tenure-track employer and a postdoc employer, differentiated based on job titles (e.g., "Assistant Professor" at CMU is labeled as employer #2, whereas "Postdoc" at CMU as #4). I treat them as separate employers because most of the postdoc jobs disappear for experienced workers at $t > 1$.

[55]Non-top firms in the industry are grouped by location. LinkedIn users often provide the location for each job. When there is no self-reported location, I use the location of a firm's headquarter as shown on Crunchbase. "West Coast (other than SF)" includes Seattle/So-Cal/Texas, and "East Coast" comprises NY/NJ/DC/New England/Chicago. If there is neither a self-reported location nor information about headquarters, I label it as "Missing".

Labor Market. At $t = 1$, workers observe job postings from all 16 employers, and draw GEV-distributed idiosyncratic preferences $\{\epsilon_{i1j}\}$ (3.2). Labor supply to an employer j is expressed as the probability of choosing j within nest $G(j)$ times the probability of choosing nest $G(j)$, as expressed in (10.1). At $t > 1$, workers actively search for new jobs with some probability $\lambda_G(\cdot)$ (3.4). A job movement is defined as $j(i, t + 1) \neq j(i, t)$.[56]

Under the assumptions above, the joint likelihood of job movements and innovation outputs is written as:

$$L(\Gamma, \theta) = ln \left(\prod_i \prod_{t=1}^{10} Pr(j_{it}, y_{it}, \widetilde{y}_{it} \mid \hat{\alpha}, \pi_1(I_{i1}, \theta), \Gamma) \right) \tag{6.2}$$

$$= \underbrace{Pr(j_{i1} \mid \pi_{i1})}_{t=1 \text{ employer}} \times \underbrace{Pr(y_{i1}, \widetilde{y}_{i1} \mid \tau_{i1}, \hat{\alpha})}_{t=1 \text{ innovation}}$$

$$\times \underbrace{Pr(j_{i2} \mid \pi_{i1}, j_{i1}, y_{i1}, \widetilde{y}_1)}_{t=2 \text{ employer} \mid \text{ public } I_{i2}, \text{ private} \widetilde{y}_{i1}} \times \underbrace{Pr(y_2, \widetilde{y}_2 \mid \tau_{i2}, \hat{\alpha})}_{t=2 \text{ innovation}}$$

$$... \times Pr(j_{10} \mid I_{10}, \widetilde{y}_9) \times Pr(y_{10}, \widetilde{y}_{10} \mid \tau_{i10}, \hat{\alpha})$$

where Γ includes all model parameters, and θ are nuisance parameters that govern the initial prior π_{i1} drawn from (6.1) given public information I_{i1}. Information $\{I_{it}\}$ evolves according to 3.5 and 3.6. Labor supply to each employer and task allocation τ are equilibrium objects solved as fixed points given model parameters Γ.

6.1.2 Parameters and Identification

Table 9 provides an overview of the parameters and indicates if they are calibrated or estimated. Identification of the model parameters relies on revealed preferences. It exploits movements between employers and variation in innovation outputs that generate differential belief updating about workers.[57]

To begin, there is a set of nuisance parameters $\{\mu_H, \sigma_H, \mu_L, \sigma_L\}$ and mixing weights $\{z_{i1}\}$

---

[56]Moving from firm A to B within the same employer (group) j is not considered as a job movement in the structural estimation. This restriction can be relaxed when I define more granular employer groups.

[57]There is variation in public belief π's at $t > 1$ as long as there is variation in innovation outputs. Suppose initially $\mu_H = \mu_L = \mu$ and $\sigma_H = \sigma_L = 0$. Prior $\pi_{i1} \equiv \mu$. As long as some workers publish while others don't, we have different posterior beliefs $\pi_{i2} \in \{P(H|\mu, y_1 = 1), P(H|\mu, y_1 = 0)\}$.

that govern the distribution of initial common belief π_{i1} as in 6.1. Different nuisance parameters generate differential priors about workers and affect the subsequent belief updating by employers, innovation production, and sorting patterns. I maximize the profile likelihood conditional on each set of nuisance parameters, and select the combination that yields the highest joint likelihood in 6.2 (see estimation procedure in 6.1.3).

The first set of model parameters in Table 9 characterizes the labor supply of workers to differentiated employers that are grouped into four nests $G \in$ {Tenure-Track, Postdoc, Top Firms, non-top Firms}. When a worker is searching for new jobs, she draws GEV-distributed idiosyncratic preferences that are independent across time and nests but can be correlated within a nest, approximately with a coefficient $1 - \rho_G$ (3.2). The nest-specific $\{\eta_{\cdot,G}\}$ allow the inclusive value of choosing a nest to depend on market belief π (10.1).[58] Workers have a preference b for (log) wages as in (3.3), which matters for the elasticity of labor supply.

The second set of parameters concerns the labor market dynamics. Workers from nest G search for new jobs with probability $\lambda_{\cdot,G}$. Academic employers are open to workers from industry with probability Λ_{JA}, and industry employers accept workers from academia with probability Λ_{AJ}. These parameters are identified from workers' movements between employers at $t > 1$. A higher $\lambda_{\cdot,G}$, for example, would predict higher turnovers for workers from nest G, and a higher $\lambda_{1,G}$ implies higher-π workers search for new jobs at higher rates. The presence of job movers and variation in π allow me to identify the labor market parameters.

The third set of parameters specifies the productivity of employers. Each employer j has a productivity f_j in routine tasks, and a proportional increase in productivity g_j in innovation tasks. $\{f_j\}$ are identified from the distribution of average workers across employers, whereas $\{g_j\}$ are identified from differential sorting and innovation outputs by workers who are more likely to be H-ability relative to those who are more likely to be L.

On the workers' side, I estimate the rates at which H vs. L can produce a research paper per unit of time on innovation, denoted by (h, l), and the rates at which they produce a higher-quality paper with a matched patent application, $(\widetilde{h}, \widetilde{l})$. In particular, $(\widetilde{h}, \widetilde{l})$ are allowed to vary between nests of employers to capture the fact that patenting is more common in industry than in academia.

---

[58]For example, $\eta_{\cdot,\text{Tenure-Track}} > 0$ would increase the share of higher-π workers on the tenure track relative to the baseline (postdocs, with $\eta_{\cdot,P} \equiv 0$). Variation in public belief π's allows us to identify such parameters.

6.1.3 Estimation Procedure

The nuisance and structural parameters are estimated in four steps.

Step 0. Given a guess of the nuisance parameters $\theta = (\mu_H, \sigma_H, \mu_L, \sigma_L)$ and the calibrated mixture weights, draw initial priors π_{i1} from 6.1. Place $\{\pi_{i1}\}$ on a discrete grid with 20 equally split intervals.

Step 1. Given a guess of model parameters

$$\Gamma = \left(\underbrace{b, \{\rho_G, \eta_{\cdot,G}\}}_{\text{Labor Supply}}, \ \underbrace{\{\lambda_{\cdot,G}\}, \Lambda_{AJ}, \Lambda_{JA}}_{\text{Labor Market Dynamics}} , \ \underbrace{\{f_j, g_j\}, \theta, \zeta}_{\text{Firm Productivity}} , \ \underbrace{h, l, \{\widetilde{h}_G, \widetilde{l}_G\}}_{\text{Worker Productivity}} \right)$$

solve for each employer's optimal contracts given public information I, or private info $\widetilde{I}$ about a worker (details in Appendix D1):

$$\left(w_{tj}^{(1)}(\widetilde{I}; \Gamma), \ \tau_{tj}^{(1)}(\widetilde{I}; \Gamma) \right) \text{ for incumbents; } \left(w_{tj}^{(0)}(I; \Gamma), \ \tau_{tj}^{(0)}(I; \Gamma) \right) \text{ for new workers}$$

that yields a fixed point in the labor supply to any (t, j), by incumbent and new workers at any possible state $(I, \widetilde{I})$:

$$p(w(\boldsymbol{p})) = \boldsymbol{p} \tag{6.3}$$

Step 2. Compute the joint likelihood of employment history and innovation outputs, given labor supply $\{p_{tj}\}$ and task allocations $\{\tau_{tj}\}$ in equilibrium. Find $\widehat{\Gamma}_\theta$ that maximize the profile likelihood conditional on nuisance parameters θ:

$$\widehat{\Gamma}_\theta = argmax_\Gamma \ L(\boldsymbol{D}; \theta, \Gamma) = log \left(\prod_{i,t} Pr(j_{it}, y_{it}, \widetilde{y}_{it} \mid \hat{\alpha}_i, \pi_{i1}, \Gamma) \right) \tag{6.4}$$

Step 3. Repeat Steps 0-2 for each guess of nuisance parameters $\theta = (\mu_H, \sigma_H, \mu_L, \sigma_L)$ on a grid. Compute the joint likelihood in a hold-out sample, $\hat{L}^{\text{test}}(\hat{\gamma}^{MLE}; \mu_H, \sigma_H, \mu_L, \sigma_L)$. Finally, find the optimal

$$\widehat{\theta} = argmax_\theta \ L(D; \theta, \widehat{\Gamma}_\theta) \tag{6.5}$$

.

6.2 Estimation Results

I estimate the structural parameters on a balanced, ten-year panel of 12,829 workers who obtained a PhD between 2000 and 2012. This sample is comparable to the full sample that I use to test for employer learning in Section 5 (see Table 2). The estimated model is able to fit the observed allocation of workers across employers, over the first decade post PhD. The estimates

suggest employers update beliefs according to workers' innovation outputs. Workers with higher employer beliefs are increasingly concentrated at top firms in industry or staying in academia, consistent with the reduced-form findings in Section 5.

Each round of estimation begins with a guess of nuisance parameters $(\mu_H, \sigma_H, \mu_L, \sigma_L)$, given which I draw a prior for each individual from (6.1) conditional on initial information. To avoid over-fitting, I estimate the model parameters by maximizing (6.4) on a 75% random sample of workers. Figure 5 shows the profile likelihood given each combination of nuisance parameters computed on the left-out 25% sample. The mixture weight on the first normal distribution centered around $\overline{\mu}$ is calibrated as the share of $\hat{H}$-ability (Appendix Table D1). Figure 5(a) shows that $\overline{\mu} > \underline{\mu}$ performs better than the lower right where $\overline{\mu} \ll \underline{\mu}$. However, based on initial information alone it is difficult to tell which workers are high-ability.[59] The optimal distributions of initial log odds in 5(a) are centered around $\overline{\mu} = -0.40$, $\underline{\mu} = -1.09$ (equivalent to the probability of being H around 0.40 or 0.25). Using a finer grid around these estimates, I find the joint likelihood (6.2) to be maximized given an initial mixture $z_{i1} \times \mathcal{N}(-0.66, 0.50) + (1 - z_{i1}) \times \mathcal{N}(-1.52, 0.20)$ as shown in Figure 5(b).

Employers Bayesian update beliefs about a worker based on innovation outputs. High-ability workers are estimated to be almost twice as likely to produce a paper per unit of time on innovation tasks as the L-ability, and three times as likely to produce a higher-quality paper with a matched patent application (Table 9).[60] Figure 6 displays the distribution of beliefs about $\hat{H}$ vs. $\hat{L}$ workers at different experience levels. At $t = 1$, beliefs about the two groups, drawn from the mixture above, overlap substantially. Once innovation takes place, the distributions quickly diverge, as evidenced by the density plots at $t = 3$ or $t = 5$. At $t = 10$, about 40% of $\hat{H}$ workers are believed to be H-ability with an odd above 3 (probability 0.95), whereas about three-quarters of $\hat{L}$ workers are believed to be H with an odd -4.5 (probability 0.01). The relative entropy between these two distributions increases from 0.03 at $t = 1$ to 2.81 at $t = 10$. Together the estimates suggest employers can learn about a worker's research ability fairly quickly from their innovation outputs post PhD.

The estimated model can fit the allocation of workers across employers and over time. The probability that a worker chooses an employer is solved iteratively as a fixed point (6.3), given a

---

[59] For example, 51-62% of workers whose first job is a tenure-track position in academia are $\hat{H}$-workers 10 years later. If $\mu_H \gg \mu_L$, employers would make a lot of mistakes.

[60] The estimates $\hat{h} > \hat{l}$ and $\hat{\overline{h}} > \hat{\overline{l}}$ in Table 9 validate Assumption 1 upon which model predictions 1, 2, 3 are derived.

guess for the model parameters. At the optimal parameters that maximize the joint likelihood of employment history and innovation outputs (6.2), the predicted share of workers at each employer falls roughly on the 45-degree line that matches with the actual shares, at different periods shown in Figure 7.

I replicate Figure 1 in simulated samples given the optimal parameters. The upward mobility by workers who are at non-top firms at some point between 1-5 years post PhD is not targeted directly in the estimation. The estimated model can fit the trajectory of $(1,1)$ workers who produce a paper with a matched patent while at non-top firms better than the other two groups (Appendix Figure D2). Despite the limitations, the model is able to capture the divergence in upward mobility between workers who start at non-top firms but produce different research signals.

The equilibrium wages, solved as a fixed point in Step 1, are non-decreasing in employer belief π. Consistent with the Proposition 1, the estimated wage returns to π are positively related to an employer's productivity in innovation, denoted by g_j. In particular, top firms are more productive in innovation (Table 10), and pay disproportionately higher wages for workers who are more likely to be H-ability, with an exception of Apple that do not let workers publish as much as other top firms.[61] Workers in academia are estimated to be less likely to move between employers than those in industry (see $\hat{\lambda}$'s in Table 9). As a result, academic employers set a flatter wage schedule, despite of being more productive in research. Additional insights such as that new workers receive a higher front-loaded wage than equally productive incumbents will be examined in future research where wages are observed.

In summary, the estimated model that maximizes the likelihood (6.2) can capture the allocation of workers across employers and over time. The profile likelihood approach allows me to find a mixture from which the initial belief about a worker is drawn. The estimates suggest employers are able to learn about a worker's ability fairly quickly from observed innovation outputs. Employer belief about $\hat{H}$ worker (top 10% based on citations) moves towards 1, whereas the belief about $\hat{L}$ moves towards 0.

---

[61]Apple Inc. appears to adopt a more secret approach. Workers at Apple rarely publish. The lack of research papers generates a low estimate of the innovation productivity at Apple (Table 10). The benchmark model cannot identify an innovative firm that are more secret than others. I will consider a firm's publication policy more formally in future work.

6.3 Counterfactual Analysis: Reducing Asymmetric Information

To quantify the impact of asymmetric employer learning on innovation productivity and talent revelation, I consider a counterfactual scenario where innovation outputs, i.e. papers and matched patent applications, become public information at the same time. I refer to the simultaneous disclosure as a scenario of symmetric information.[62]

Given the estimates in Table 9, I forward simulate the employment path and research production by workers in the balanced sample, holding fixed initial information.[63] For the counterfactual, I begin with the same set of workers with the same initial information and prior π_{i1} drawn from (6.1). I find each employer's optimal wages and task allocations under symmetric information disclosure, which are solved as a fixed point as in 6.3. The key difference is that employers have equal access to any innovation output in the past in the counterfactual. I repeat the simulation above multiple times. The average change in each equilibrium outcome, such as job mobility, innovation productivity or the revelation of H-ability workers, is interpreted as an estimate of the impact of asymmetric learning.

Job Mobility. Figure 8 shows the gap in next-year job mobility between $(1,0)$ and $(1,1)$ workers, which has been tested formally via regression 5.1 (Table 5). Under asymmetric information, $(1,1)$ who produce a paper with a matched patent application are less likely to move than $(1,0)$ who have a paper only. If patents are disclosed without delay, as shown in the pink bars (right), $(1,1)$ workers at non-top firms are 1.6 ppt more likely to move into a new employer the next year. The difference-in-difference in any job mobility from non-top firms is comparable to but slightly lower than the regression estimate $\hat{\gamma}_{11} = 0.018$ on lagged innovation signal $(1,1)$, denoted by $D_{it}(1,1)$ in 5.1. Consistent with the reduced-form evidence in Table 5, the effects are the largest for workers who are employed by non-top firms, which could not increase wages as much as top firms for workers who are revealed to be more productive according to the estimates (Appendix Figure D3).

Upward mobility from non-top firms to top firms also increases by 1.6 ppt as information

---

[62]There may well be productive signals that remain private at an incumbent employer. This paper, however, focuses on the observable margin of asymmetric information where a matched patent application becomes public information with some delay. The estimated impacts of asymmetric information may be interpreted as a lower bound of the role of information frictions between employers.

[63]Initial information I_{i1} includes whether person i graduates from one of the top 25 CS programs, whether her first job is research related (based on job title), and the initial nest of her first job $G(i, 1) \in \{$Tenure-track, Postdoc, Top Firms, non-top Firms$\}$. Mixture weights in 6.1 are calibrated conditional on the initial information. In each simulation, I draw new priors $\{\pi_{i1}\}$ but hold fixed the initial information and mixture weight z_{i1}.

becomes more symmetric, as shown by Figure 8(b). The diff-in-diff counterfactual estimate is comparable to but slightly higher than the regression estimate $\hat{\gamma}_{11} = 0.013$. Focusing on workers from non-top firms, I compare the upward mobility between workers who produce $(1,1)$, $(1,0)$ and $(0,0)$ at non-top firms at some point in the first five years after PhD. Figure 9 shows that simultaneous information disclosure does not change the upward mobility by $(0,0)$ workers, which makes sense as they are not affected by the policy. In contrast, $(1,0)$ and $(1,1)$ workers are increasingly sorted into top firms, and the increase is larger for $(1,1)$ who directly benefit from the increase in public information about their research ability.[64]

Talent Revelation. Reducing asymmetric information increases the chance that a H-ability worker is publicly revealed. I define talent revelation as the fraction of H-ability workers who have a public belief ranked among the top 10% across all workers, at a given period. Since I hold fixed the initial priors $\{\pi_{i1}\}$ drawn from 6.1, there is no difference in talent revelation at $t = 1$ under asymmetric vs. symmetric information. After the first period, talent revelation would increases 2-4 ppt's or 4-6% higher when information were more symmetric (Table 11).

Innovation Productivity. The counterfactual analysis provides an answer to whether asymmetric learning hurts innovation productivity. First, note the estimated model can closely match the average innovation rates in the data. On average 9.4% of workers in the sample produce a paper each period, and 1.7% produce a paper with a matched patent application, both of which are matched by the simulations under asymmetric learning (first column of Table 11). If matched patents are disclosed at the same time as papers, 9.9% of workers are expected to produce a paper each period, which is a 4.7% increase relative to the asymmetric benchmark, as shown in Table 11. In addition, there is 6.2% increase in the rate at which workers produce a higher-quality paper with a matched patent.

The increase in innovation productivity under symmetric information can be explained by a combination of more positive assortative matching and that employers could allocate new recruits more efficiently without the delayed signal. The 6% increase in patenting is comparable to the increase if a worker's true ability $\alpha_i \in \{H, L\}$ is fully revealed to all employers at $t = 8$ or $t = 9$ (Appendix Figure D4). The changes in innovation under symmetric information can be translated

---

[64]$(1,0)$ workers from non-top firms are still more likely to be H-ability than $(0,0)$ workers. They may produce papers with a matched patent elsewhere and are recognized as high-ability later on. In other words, $(1,0)$ workers from non-top firms may indirectly benefit from the reduction of asymmetric information when they stay productive.

to 650 more papers and 130 more papers with a matched patent, by thirteen thousand workers in this panel across the first decade of their career. One limitation of this model is that it ignores the heterogeneity in the quality of these works. Given the considerable skewness in the distribution of innovation value, one of the 650 papers or 130 matched patents could be extremely impactful (Gambardella, Harhoff, and Verspagen 2010; Hall and Harhoff 2012).

Another missing piece of the model is that firms may want to delay publishing papers if matched patents were to be revealed simultaneously, an action that would have attenuated the increase in innovation due to talent revelation. It is unclear if firms would do that, however, given the fast pace of CS research especially in AI-related fields and higher returns to be the first to publish (e.g., Hill and Stein 2022).[65]

7 Conclusion

This paper presents novel evidence of asymmetric employer learning in the labor market for computer scientists, which is high-skilled and intensive in innovation. Finding direct evidence for asymmetric learning has been difficult, as we rarely observe private information about workers that only the incumbent employers can access. In this context, I exploit the delayed disclosure of patent applications that are matched to research papers. Employers (in industry) often seek exclusive rights to valuable innovation by filing a patent application for inventions publicized in a conference paper. The matched patent applications are disclosed more than one year later than the original papers. I build this institutional feature into a dynamic model of learning where firms make strategic decisions about how much time workers can spend on innovation, taking into account the risk of turnover when a worker is publicly revealed to be productive. This model generates predictions that I use as tests for employer learning.

I find that job mobility increases within a year for workers who produce any new paper, and the increases are strongest among those working outside the top tech firms. Before the revelation of a patent, workers with a new paper and matched patent move less than coworkers with a paper only. After the revelation, however, they are 13% more likely to move out of a non-top firm and 35% more likely to move into a top firm. The delayed mobility responses among authors on papers with

---

[65]I will consider a more generalized framework that takes into account firms' competition in publishing and study its implications on workers.

a matched patent provide evidence of asymmetric learning.

It is unclear from the empirical tests whether total innovation would increase if the asymmetry is removed. To answer that question, I estimate the model and consider the counterfactual where the matched patents are disclosed simultaneously as papers. Under the counterfactual, the annual upward mobility of workers with a patent from non-top to top firms would increase by 32%, comparable with the reduced-form estimate. Total innovation outputs increase by about 5%, which is driven by an increase in positive assortative matching and the fact that employers can allocate new recruits more efficiently given more information.

One limitation about the data is that CS Ph.D.'s on LinkedIn are more likely to work in industry than in academia. Most of the empirical analysis in this paper focuses on workers in industry, and in particular their transitions from non-top to top firms as an indicator for upward job mobility. It may be interesting to look further into the AI brain drain issue that tech firms are poaching talent from academia (e.g., Jurowetzki, Hain, Mateos-Garcia and Stathoulopoulos 2021). In future work, I aim to enrich the data with CVs of faculty in computer science, which will give a more complete picture of the CS labor market. Building on the model in this paper, I plan to investigate if increasing academic jobs for Ph.D. workers in industry could counter the issues of brain drain, and also increase the total revelation of talent.

Another important research question is whether workers know more about themselves, whether they are aware of the information asymmetry, and how they might signal their ability to outside employers to reduce the information frictions. This paper focuses on the decisions by employers. In other works, I let workers be forward-looking and study their own signaling through open-source contributions, for example. A survey on the computer scientists would also help answering these questions.

Asymmetric information is prevalent across labor markets. Ph.D. computer scientists may have already been the privileged to be able to publish. It will be helpful to study the impacts of asymmetric employer learning in various labor markets, and consider policies that could help reduce the information frictions between employers.

References

Acemoglu, Daron, and Jorn-Steffen Pischke. 1998. "Why Do Firms Train? Theory and Evidence." *The Quarterly Journal of Economics*, 113(1): 79–119.

Ahmadpoor, Mohammad, and Benjamin F. Jones. 2017. "The dual frontier: Patented inventions and prior scientific advance." *Science*, 357(6351): 583–587.

Altonji, Joseph G., and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *The Quarterly Journal of Economics*, 116(1): 313–350.

America Invents Act, 35 U.S.C. 102. 2011*a*. "Conditions for Patentability; Novelty."

America Invents Act, 35 U.S.C. 122. 2011*b*. "Confidential Status of Applications; Publication of Patent Applications."

American Inventors Protection Act, 35 U.S.C. 1999.

Ashenfelter, Orley, and Ruth Gilgenbach. 2023. "No-Poaching Agreements as Antitrust Violations: Animation Workers Antitrust Litigation." *Princeton IRS Working Paper 657*.

Ashenfelter, Orley, David Card, Henry Farber, and Michael R Ransom. 2022. "Monopsony in the Labor Market: New Empirical Results and New Public Policies." *Journal of Human Resources*, 57(3): S1–S10.

Becker, Gary. 1964. "Human Capital." University of Chicago Press.

Bernhardt, Dan. 1995. "Strategic Promotion and Compensation." *Review of Economic Studies*, 62(2): 315–339.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1999. "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy." *American Economic Review*, 89(3): 400–430.

Boozer, Michael. 1994. *Wages and Turnover with Imperfect Information and Imperfect Competition.* Doctoral Dissertation, Princeton University.

Bresnahan, Timothy. 1981. "Departures from marginal-cost pricing in the American automobile industry: Estimates for 1977-1978." *Journal of Econometrics*, 17(2): 201–227.

Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review*, 39(2): 257–273.

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.

Cederlöf, Jonas, Arash Nekoei Peter Fredriksson, and David Seim. 2023. "Mandatory Notice of Layoff, Job Search, and Efficiency." *Working Paper*.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *The American Economic Review*, 67(3): 297–308.

Farber, Henry S., and Robert Gibbons. 1996. "Learning and Wage Dynamics." *The Quarterly Journal of Economics*, 111(4): 1007–1047.

Ferreira, Daniel, and Radoslawa Nikolowa. 2022. "Talent Discovery and Poaching Under Asymmetric Information." *The Economic Journal*, 133(649): 201–234.

Gambardella, Alfonso, Dietmar Harhoff, and Bart Verspagen. n.d.. "The value of European patents." *European Management Review*, 5(2): 69–84.

Gibbons, Robert, and Lawrence Katz. 1991. "Layoffs and Lemons." *Journal of Labor Economics*, 9(4): 351–380.

Gibbons, Robert, and Lawrence Katz. 1992. "Does Unmeasured Ability Explain Inter-Industry Wage Differentials." *The Review of Economic Studies*, 59(3): 515–535.

Gibson, Matthew. 2023. "Employer market power in Silicon Valley." *Working Paper*.

Goldberg, Pinelopi Koujianou. 1995. "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry." *Econometrica*, 63(4): 891–951.

Graham, Stuart J. H., Alan C. Marco, and Amanda F. Myers. 2018. "Patent transactions in the marketplace: Lessons from the USPTO Patent Assignment Dataset." *Journal of Economics & Management Strategy*, 27(3): 343–371.

Greenwald, Bruce C. 1986. "Adverse Selection in the Labour Market." *Review of Economic Studies*, 53(3): 325–347.

Haegele, Ingrid. 2022. "Talent hoarding in organizations." *arXiv 2206.15098*.

Hall, Bronwyn H, and Dietmar Harhoff. 2012. "Recent Research on the Economics of Patents." *Annual Review of Economics*, 4(1): 541–565.

Hendricks, Kenneth, and Robert H Porter. 1988. "An Empirical Study of an Auction with Asymmetric Information." *The American Economic Review*, 865–883.

Hill, Ryan, and Carolyn Stein. 2022. "Scooped! Estimating Rewards for Priority in Science." *Working Paper*.

Imbens, Guido, and Jeffrey Wooldridge. 2007. *What's New in Econometrics Lecture 11: Discrete Choice Models.* NBER.

Jurowetzki, Roman, Daniel Hain, Juan Mateos-Garcia, and Konstantinos Stathoulopoulos. 2021. "The Privatization of AI Research(-ers): Causes and Potential Consequences - From university-industry interaction to public research brain-drain?" *arXiv 2102.01648*.

Kahn, 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.

Kirpalani, Rishabh, and Erik Madsen. 2023. "Strategic investment evaluation." *Theoretical Economics*, 18(3): 1141–1180.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar. 2019. "Who Profits from Patents? Rent-Sharing at Innovative Firms." *The Quarterly Journal of Economics*, 134(3): 1343–1404.

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

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

Magerman, Tom, Bart Van Looy, and Koenraad Debackere. 2015. "Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology." *Research Policy*, 44(9): 1702–1713.

Manning, Alan. 2003. "Human Capital and Training." *Monopsony in Motion: Imperfect Competition in Labor Markets*, 301–322. Princeton University Press.

Manning, Alan. 2012. "Steady-state Equilibrium in a Model of Short-Term Wage-Posting." *CEP Discussion Paper No 1158*.

McFadden, Daniel. 1973. "Conditional logit analysis of qualitative choice behavior." In *Fontiers in Econometrics*. , ed. Paul Zarembka, 105–142. New York:Academic press.

Morrow, W. Ross, and Steven J. Skerlos. 2011. "Fixed-Point Approaches to Computing Bertrand-Nash Equilibrium Prices Under Mixed-Logit Demand." *Operations Research*, 59(2): 328–345.

Neelakantan, Arvind, Tao Xu, Raul Puri, Alec Radford, Jesse Michael Han, Jerry Tworek, Qiming Yuan, and et al. 2022. "Text and Code Embeddings by Contrastive Pre-Training."

Pallais, Amanda. 2014. "Inefficient Hiring in Entry-Level Labor Markets." *American Economic Review*, 104(11): 3565–99.

Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica*, 70(6): 2295–2350.

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

Stern, Scott. 2004. "Do Scientists Pay to Be Scientists?" *Management Science*, 50(6): 835–853.

Stevens, Margaret. 1994. "A Theoretical Model of On-the-Job Training with Imperfect Competition." *Oxford Economic Papers*, 46(4): 537–62.

Strobl, Günter, and Edward D. Van Wesep. 2013. "Publicizing Performance." *Management Science*, 59(4): 918–932.

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

US, Department of Justice. 2010. "Complaint, United States of America v. Adobe Systems."

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

8 Figures

Figure 1: Different Paths of Employment by Top Firms*, by Earlier Research Outputs at non-top Firms
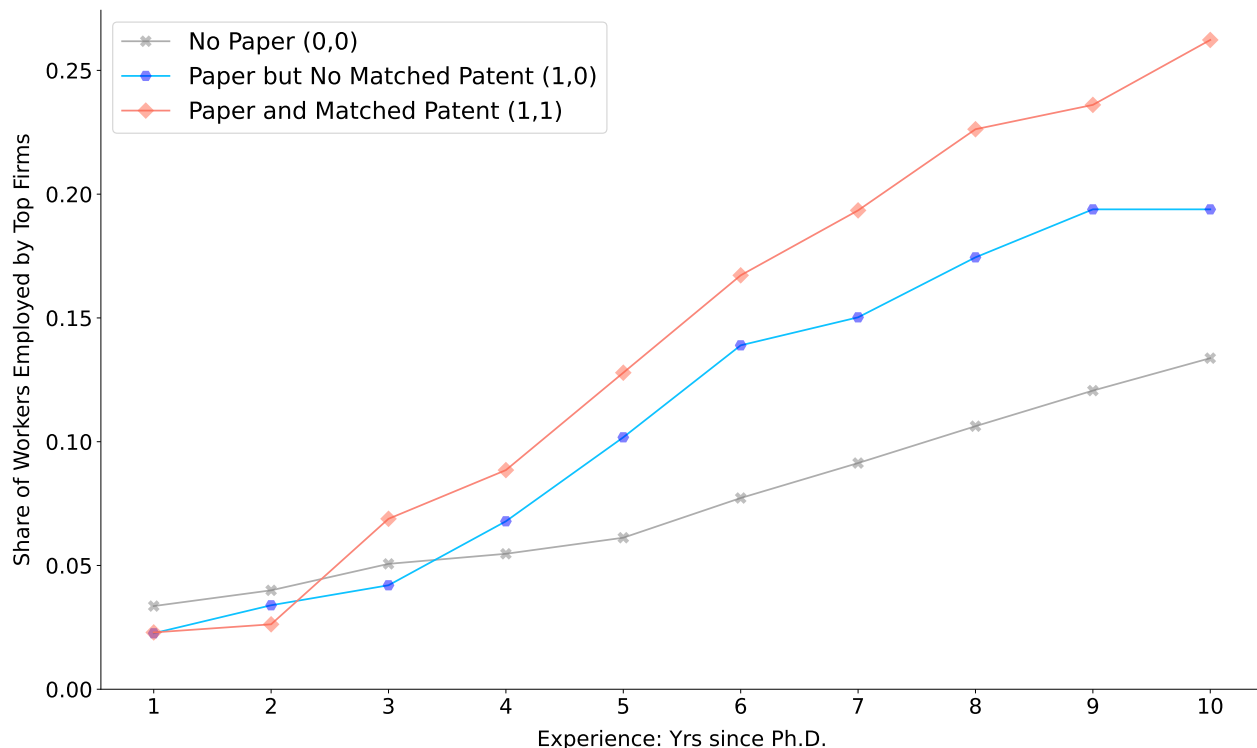
Notes: This figure shows the share of workers who are employed by one of the top firms in each year post PhD, for three groups of workers who are outside the top firms at some point in the first five years but produce different research outputs. The sample comprises CS/EE Ph.D. workers with nonmissing full-time employment records throughout the first decade post PhD. The $(0,0)$ group refers to workers who do not produce any CS publication while being employed by a non-top firm. The $(1,0)$ group produces at least one CS paper at non-top firms but none of the papers have a matched patent application. The $(1,1)$ group produces CS paper(s) and at least one of the papers can be matched to a patent application, which often indicates more influential research.

* Throughout the paper, top firms refer to {Google, Microsoft, IBM, Facebook, Amazon, Apple}. non-top firms refer to other industry employers.

Figure 2: Citations over time: Papers With and Without a Matched Patent Application

Citation Path: Papers with a Matched Patent vs. without
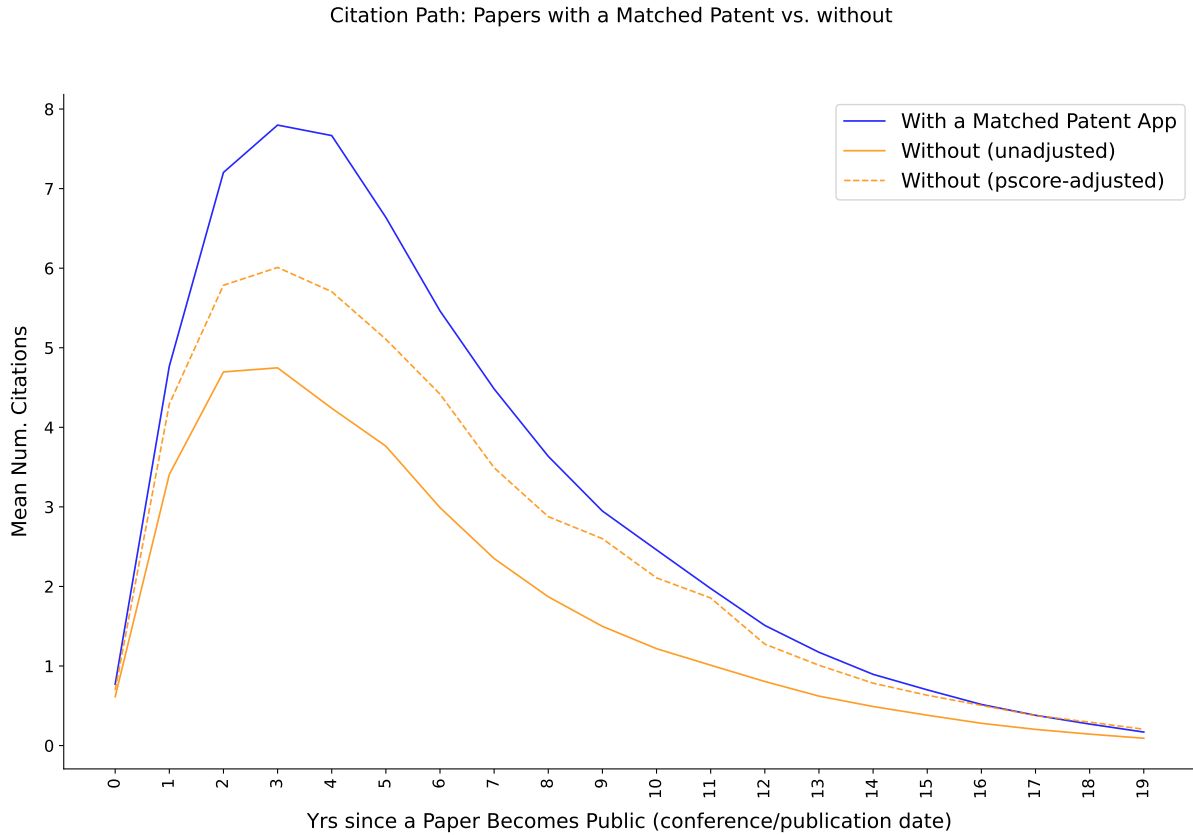
Notes: This figure shows the mean citations received by CS research papers by year since the paper becomes public. I use citation data from Scopus, which includes citations by conference proceedings. I exclude self-citations by authors of the original paper. The blue line (top) shows the citation path for papers that are matched to a patent application filed around the same time (see Section 4.3.2). The solid yellow line (bottom) shows the path for papers without a matched patent. To take into account differences in patenting across time or employers, I estimate a Poisson regression of having a matched patent on firm (author affiliation) and year fixed effects. Following DiNardo, Fortin and Lemieux (1996), I reweight papers without a matched patent by the predicted odds of having one. The dashed yellow line in the middle shows the weighted average of citations among papers without a patent. The gap between the two groups of papers is almost 0 in the first year since publication after the propensity-score reweighting. The gap starts to expand around two years after the paper becomes public, which coincides with the disclosure of patent applications by the U.S. patents office.

Figure 3: Mobility Gaps, With vs. Without a Paper, by Group of Employers

(a) Any Move between Employers

(b) Employment by Top Firms Next Year
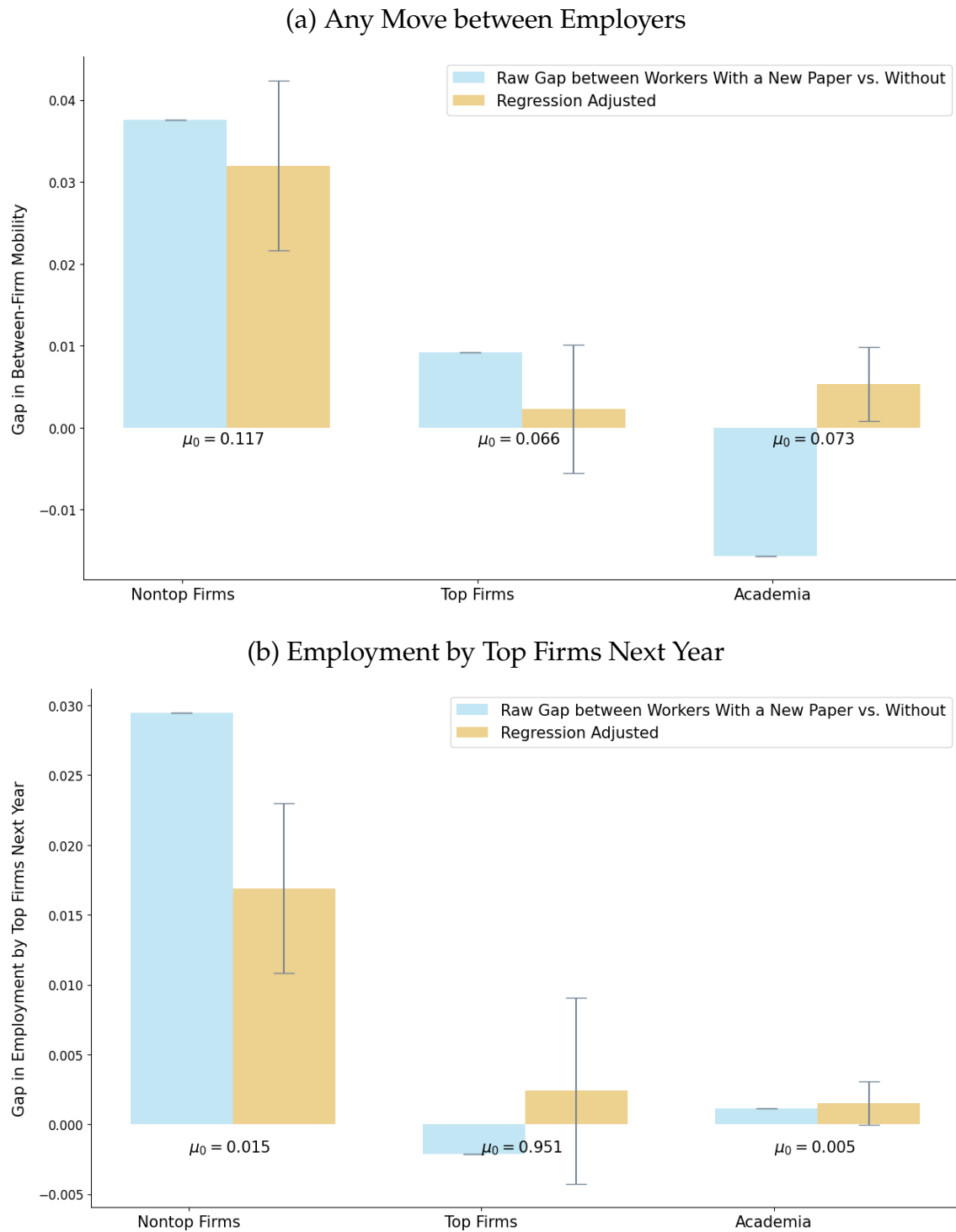
Notes: This figure shows the difference in next-year job mobility between workers who have a new CS paper this year and workers without a new paper, at non-top firms, top firms and academia, respectively. The blue bars are unadjusted raw gaps in job mobility, whereas the yellow bars are adjusted in a regression that controls for Ph.D. school, experience since Ph.D., firm-year fixed effects and other controls listed under Table 5. μ_0 refers to the mean mobility among workers without a new CS paper.

Figure 4: New PhDs and Matched Profiles by Graduation Year
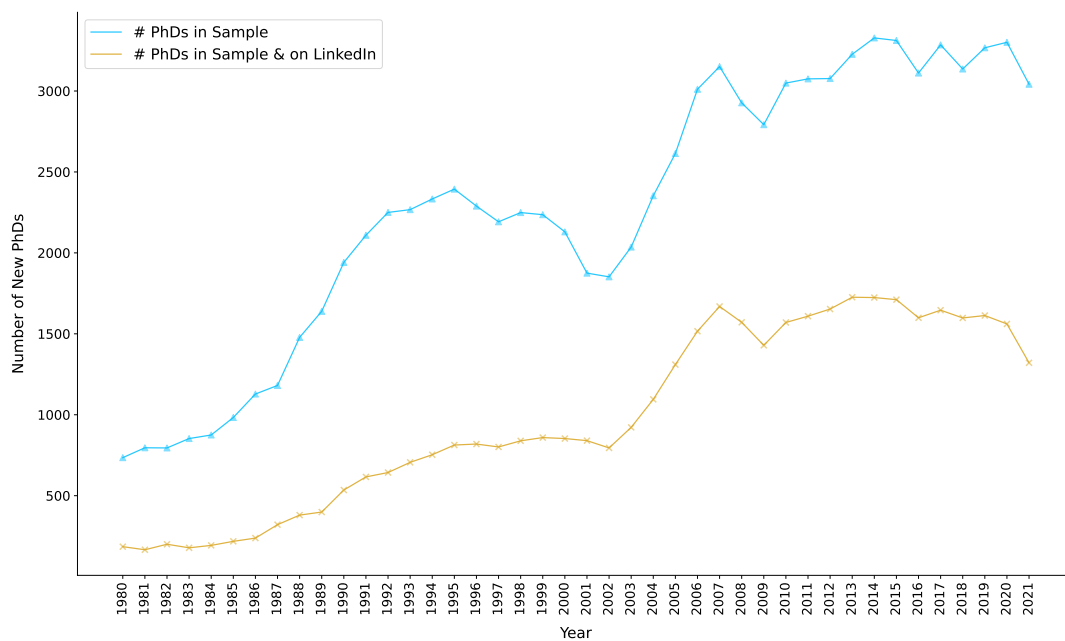
Notes: The blue line (top) shows the number of Ph.D. recipients in Computer Science or Electrical Engineering identified in ProQuest dissertation database or various school-specific sources (Appendix Table B2) by graduation year from 1980 to 2021. The yellow line plots the number of Ph.D.s who are matched with a public LinkedIn profile by full name, Ph.D. institution, year of graduation.

Figure 5: Profile Likelihood under Different Nuisance Parameters

(a)

(b)
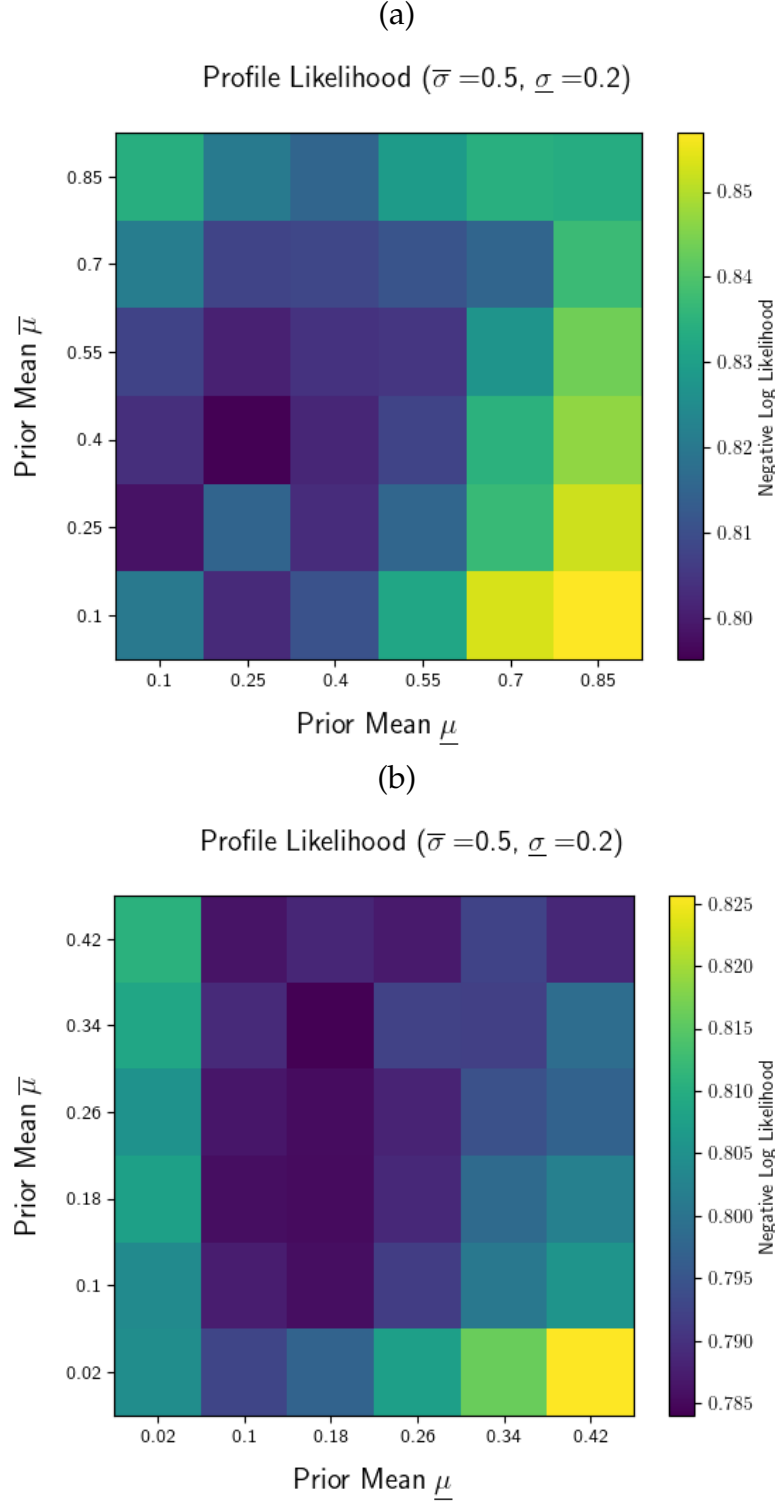
Notes: This figure shows the maximized profile likelihood (6.2, 6.4) conditional on nuisance parameters $\theta = (\overline{\mu}, 0.5, \underline{\mu}, 0.2)$. On both axes, I label the probability of being H rather than log odds. (a) shows the profile likelihood at $(0.25, 0.4)$. (b) increases the granularity by zooming into the $[0.02, 0.42]$ range and finds the optimal means to be $(0.18, 0.34)$. That is, the log odds of initial priors are drawn from either $N(-0.66, 0.50)$ or $N(-1.52, 0.20)$. Profile likelihood under alternative σ's can be found in Appendix Figure D1.

Figure 6: Distribution of Employer Beliefs (Log Odds): $\hat{H}$-workers vs. $\hat{L}$
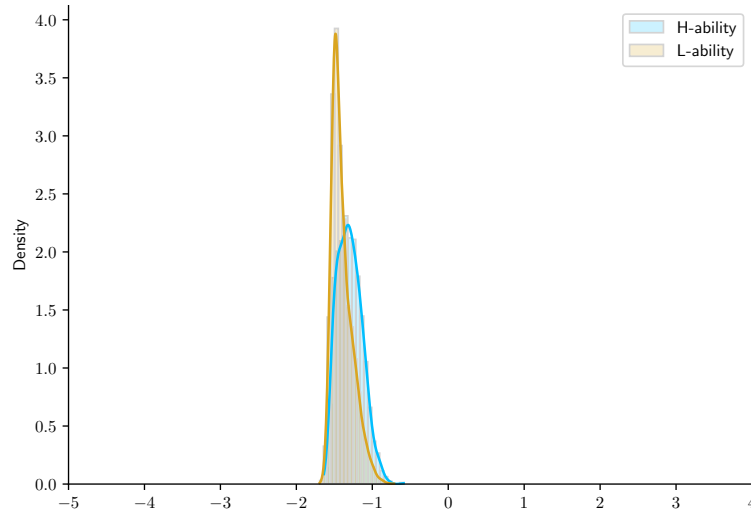
$t = 1$: Log Odds of Being H
Initial Mixture: N(-0.66, 0.5) or N(-1.52, 0.2); Rel. Entropy (Kullback-Leibler) at t : 0.031
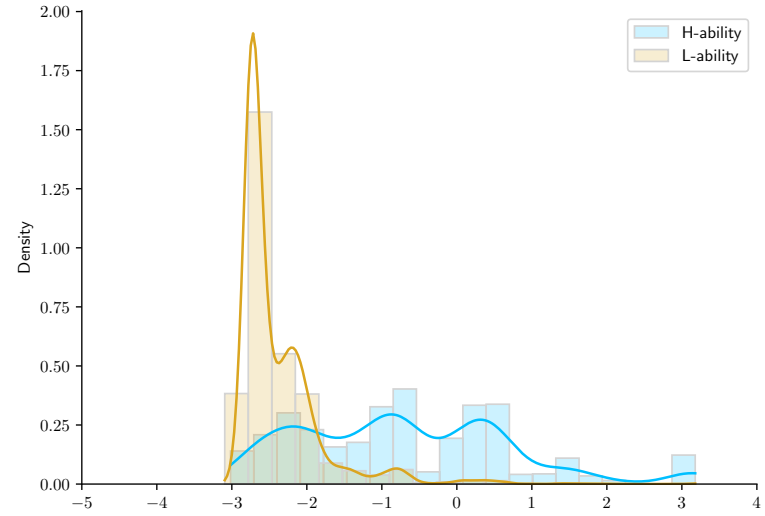
$t = 3$: Log Odds of Being H
Initial Mixture: N(-0.66, 0.5) or N(-1.52, 0.2); Rel. Entropy (Kullback-Leibler) at t : 1.399
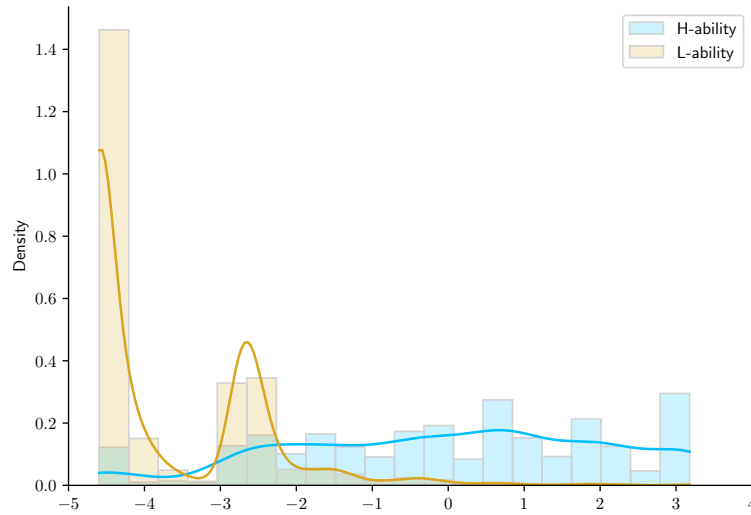
$t = 5$: Log Odds of Being H
Initial Mixture: N(-0.66, 0.5) or N(-1.52, 0.2); Rel. Entropy (Kullback-Leibler) at t : 2.191
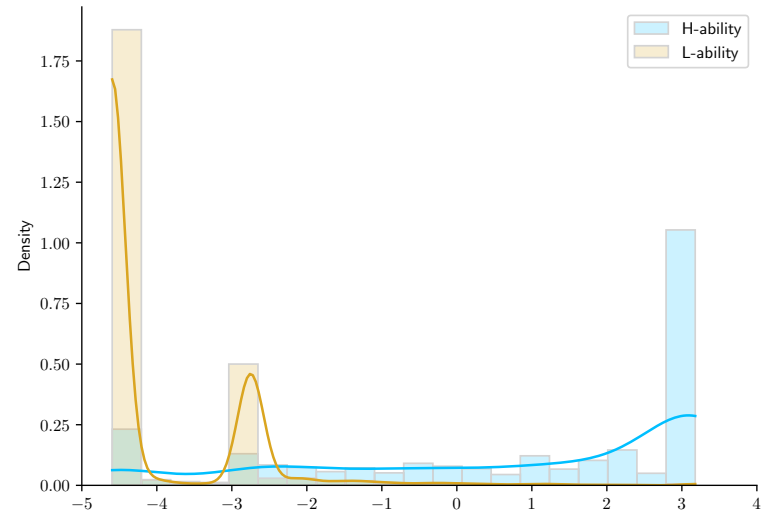
$t = 10$: Log Odds of Being H
Initial Mixture: N(-0.66, 0.5) or N(-1.52, 0.2); Rel. Entropy (Kullback-Leibler) at t : 2.805

Notes: This figure shows the distribution of employer beliefs (in log odds) at $t \in \{1, 3, 5, 10\}$, for workers who are labeled as $\hat{H}$ vs. $\hat{L}$ based on cumulative citations (see Section 6.1.1). Initial priors are drawn from $N(-0.66, 0.50)$ or $N(-1.52, 0.20)$, the optimal nuisance parameters in Figure 5.

Figure 7: Model Fit: Allocation of Workers across Employers, $\hat{p}_{tj}$ vs. p_{tj}

(a) $t = 1$ (b) $t = 3$ (c) $t = 5$ (d) $t = 10$
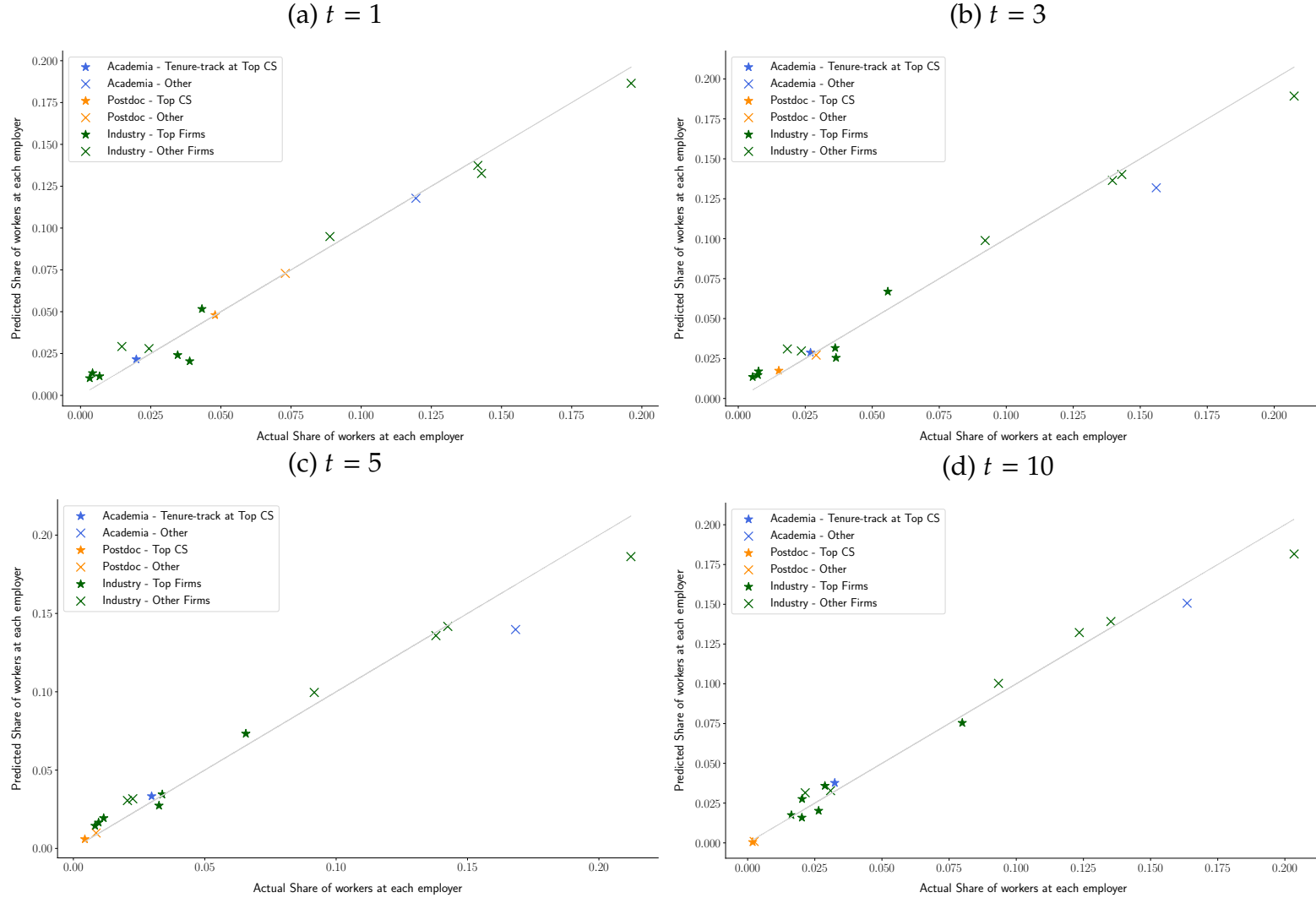
Notes: This figure shows the predicted share of workers at each employer (group) $\hat{p}_{tj}$ against the actual share p_{tj}, at different experience levels. Given the estimated parameters in Table 9, I forward simulate the employment path and innovation outputs by each worker in the balanced sample, holding fixed initial information including the initial nest at $t = 1$ (see 6.1). In the simulated sample, I compute $\hat{p}_{tj}$ as the share of workers employed by j, at experience t (yrs after PhD).

Figure 8: Differences in Mobility between $(1,0)$ and $(1,1)$

(a) Any Move, $(1,1)$ minus $(1,0)$
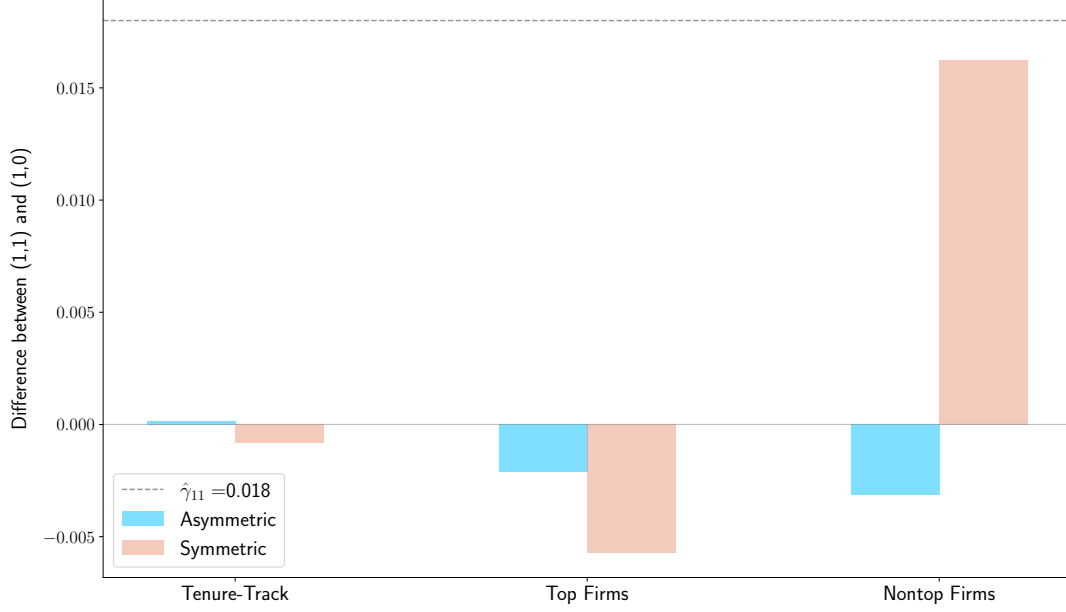
(b) Move into Top Firms, $(1,1)$ minus $(1,0)$
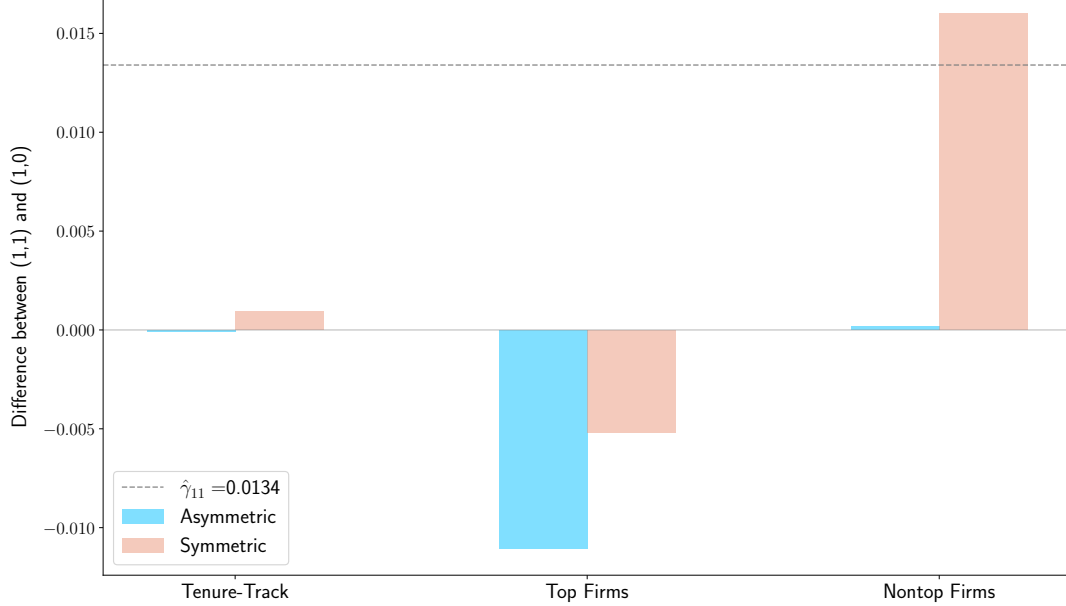
Notes: This figure shows the gap in next-year job mobility between $(1,1)$ workers who produce a paper with a matched patent and $(1,0)$ workers who produce a paper only, in forward-simulated samples given parameters in Table 9. The blue bars (left) display the gaps when the matched patent $\widetilde{y} = 1$ becomes public information with a one-year delay, whereas the pink bars (right) display the gaps when the $\widetilde{y} = 1$ is disclosed simultaneously as the paper. See text in Section 6.3 for details on the counterfactual simulation. For each mobility outcome, I show the estimated $\hat{\gamma}_{11}$ on lagged $(1,1)$ output in the regression for workers from non-top firms (Table 5, regression 5.1).

Figure 9: Upward Mobility from non-top to Top Firms, Asymmetric vs. Symmetric
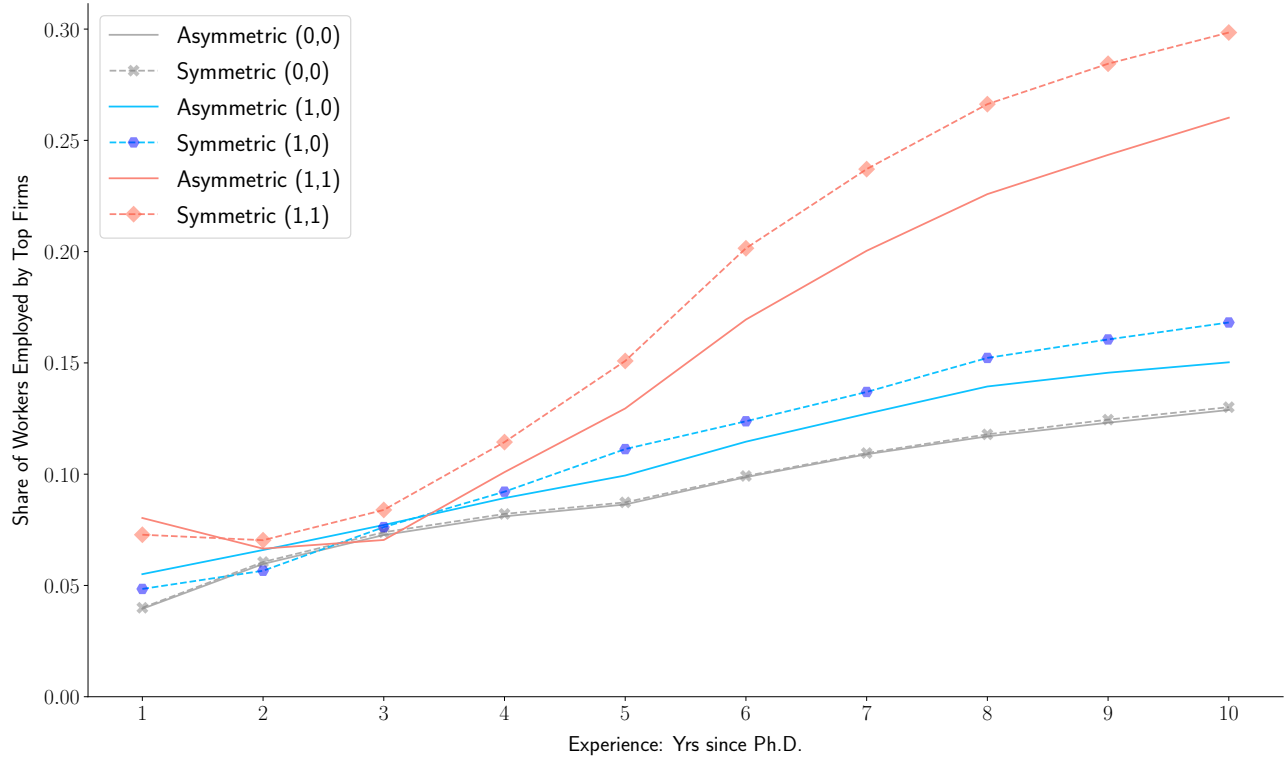
Notes: This figure replicates Figure 1 to show the upward mobility of workers who are employed by non-top firms but produce different research outputs during the first five years after PhD. The solid lines show the upward mobility for workers under asymmetric information, where matched patents become public information a year later than papers. The dashed lines show the counterfactual upward mobility if matched patents are disclosed simultaneously as papers. There is little difference for the $(0,0)$ group who do not have a paper while at non-top firms initially. In contrast, the upward mobility of workers from other groups increases, and more for the $(1,1)$ workers who can be told apart from $(1,0)$ faster under the counterfactual of simultaneous information disclosure.

9 Tables

Table 1: Examples of CS Papers and Matched Patent Applications

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

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

Table 2: Descriptive Statistics: Matched Computer Scientists

	Full Sample		Balanced sample	
	Mean	SD	Mean	SD
Gender from Name or Picture				
Female	0.118	0.323	0.103	0.304
Male	0.726	0.446	0.695	0.460
Education				
Year of Ph.D.	2006.8	9.842	2007.0	3.489
Ph.D. in CS (∋ EECS)	0.530	0.499	0.523	0.499
Ph.D. in EE	0.470	0.499	0.477	0.499
Bachelor from Top 60 CS in the U.S.	0.278	0.448	0.248	0.432
Bachelor from Top 20 Universities in the U.S.	0.131	0.337	0.122	0.327
Research Outputs Post Ph.D.				
Num. Papers	2.353	9.050	3.149	10.438
Num. Papers with a Matched Patent	0.459	2.979	0.695	3.311
Num. Papers with a Matched Patent (High-Quality)	0.225	1.544	0.360	1.826
Any Paper Post PhD	0.275	0.447	0.322	0.467
Any Paper with a Matched Patent	0.090	0.286	0.132	0.338
Any Papers with a Matched Patent (High-Quality)	0.065	0.247	0.098	0.297
Employers				
Num. Yrs with Full-time Employment	13.486	6.905	15.724	3.487
Num. Tenure-track Employers	0.285	0.589	0.299	0.604
Num. Postdoc Employers	0.153	0.396	0.163	0.410
Num. Top Firms	0.299	0.550	0.377	0.623
Num. Firms Outside the Top	1.770	1.619	1.753	1.462
Ever on the Tenure-track	0.224	0.417	0.230	0.421
Ever a Postdoc	0.140	0.347	0.149	0.356
Ever at Top Firms	0.257	0.437	0.308	0.462
Ever at Firms Outside the Top	0.778	0.416	0.792	0.406
Observations	40,118		12,446	

Notes: This table summarizes the sample of matched Ph.D.'s with non-missing full-time employment records on LinkedIn. Section 4.1-4.2 discuss the matching between Ph.D. dissertations and public LinkedIn profiles in detail. The full sample (first two columns) includes matched CS/EE Ph.D.'s from top 60 CS schools who graduated between 1980 and 2021, and have at least one full-time job with a non-missing employer listed on one's LinkedIn profile. I use the full sample throughout Section 5. The balanced sample in the last two columns is restricted to those who graduated between 2000 and 2012 and have 10 years of nonmissing job history since Ph.D. on LinkedIn. I use this subsample to estimate the 10-period structural model in Section 6.

• Gender is classified based on either first name or profile picture (available for 78% of the sample). 15% remains missing, due to either a missing picture or gender-neutral or foreign names that cannot be classified based on the U.S. Census.

• High-quality matched patent refers to those with a similar abstract as the paper's (distance in embedding ≤ 0.35).

Table 3: Descriptive Statistics: Person-Year Panel

$j(i,t) \in$	Nontop Firms		Top Firms		Academia	
	Mean	SD	Mean	SD	Mean	SD
Experience (yrs since Ph.D.)	11.640	8.552	9.239	7.344	11.648	9.056
Experience in Academia	1.098	3.028	0.697	2.264	9.803	8.176
Tenure	5.165	5.574	4.987	5.362	7.679	7.677
Current Position						
Postdoc	0.000	0.000	0.000	0.000	0.101	0.302
Research Scientist	0.119	0.324	0.148	0.355	0.037	0.190
Engineer	0.455	0.498	0.602	0.489	0.042	0.200
Manager	0.153	0.360	0.195	0.396	0.017	0.130
Senior Role	0.495	0.500	0.393	0.488	0.058	0.233
Any Promotion	0.063	0.242	0.065	0.247	0.057	0.232
Research Outputs						
Any New Paper	0.023	0.150	0.103	0.304	0.184	0.387
Any New Paper with a Matched Patent	0.009	0.095	0.052	0.222	0.022	0.146
Any New Paper with a Matched Patent (High Quality)	0.006	0.075	0.034	0.181	0.012	0.108
Movements between Employers $j(i,t)$ vs. $j(i,t+1)$						
New Employer Next Year	0.118	0.322	0.067	0.250	0.070	0.255
Employed by Top Firms Next Year	0.016	0.126	0.951	0.217	0.005	0.072
Observations	326421		69754		144845	

Notes: This table summarizes the person×year level panel for matched Ph.D.'s (full sample in Table 2). The first two columns display the means across person×year observations for those currently employed by a firm outside the top tier in the industry, denoted as $j(i,t) \in$ non-top. The second set restricts to those working at top firms, and the third set to those working in academia (including postdocs, tenure-track jobs or other roles). I put all postdocs and faculty in the third group. There are 530 person×year observations (226 individuals) where a person works as a postdoc or visiting scholar in one of the top firms.

Table 4: Papers with a Matched Patent Application vs. without

	non-top Firms		Top Firms		Academia	
Matched to a Patent?	0	1	0	1	0	1
Num. Papers	8,479	2,458	9,299	3,631	69,623	3,330
Num. Authors	6,222	2,058	3,515	1,506	14,344	2,025
Citations						
1-yr Citations	4.87	6.27	7.46	7.64	4.49	5.37
3-yr Citations cite3	15.29	23.23	23.44	29.21	14.99	19.43
5-yr Citations	23.13	36.30	35.94	51.36	23.77	32.31
Match Quality						
Text Distance	.	0.25	.	0.25	.	0.26
Team Overlap	.	0.87	.	0.85	.	0.86

Notes: This table compares CS conference papers without (0) vs. with (1) a matched patent application. A paper is considered from academia if at least 50% of matched authors are employed in academia, or from top tech firms if at least 50% are from {Google, Microsoft, IBM, Facebook, Amazon, Apple}, or from non-top firms if neither case applies. "Team Overlap", as in Table 1, is defined as the fraction of inventors on a patent application who are matched with authors on the paper. "Text distance" is measured by the distance between the embedded vector for a paper's title and abstract, and that of a patent's.

Table 5: Effects of Publications & Matched Patents on Job Mobility

	Move between Firms			Move into Top Firms		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
(Any Pub, Any Patented Pub) at t						
$d_{it}(1,0)$	0.0351	0.0030	0.0056	0.0175	0.0020	0.0014
	(0.0059)	(0.0047)	(0.0023)	(0.0034)	(0.0038)	(0.0008)
$d_{it}(1,1)$	0.0176	-0.0004	0.0046	0.0129	0.0022	0.0021
	(0.0102)	(0.0061)	(0.0064)	(0.0061)	(0.0052)	(0.0022)
(Any Pub, Any Patented Pub) between $[t-3, t-1]$						
$D_{it}(1,0)$	0.0019	-0.0004	0.0078	0.0053	0.0013	0.0034
	(0.0036)	(0.0032)	(0.0021)	(0.0018)	(0.0029)	(0.0008)
$D_{it}(1,1)$	0.0180	0.0034	0.0012	0.0134	0.0041	0.0039
	(0.0068)	(0.0051)	(0.0044)	(0.0041)	(0.0044)	(0.0017)
Mean	0.1177	0.0671	0.0702	0.0160	.9505832	.0052813
Mean|No Pub	0.1168	0.0662	0.0731	0.0153	.9508	0.0051
Mean|Any Pub	0.1544	0.0754	0.0574	0.0448	0.9487	0.0062
N	222,104	62,145	121,116	222,153	62,147	121,124
Adjusted R2	0.1034	0.0157	0.1068	0.0298	0.0108	-.0329

Notes: This table presents regression estimates of equation 5.1. The estimation sample is at Person $\times$ Year level, restricted to years with non-missing full-time employment after PhD. The first three columns show the results for any move between firms as the dependent variable, $M_{it} = 1[j(i, t+1) \neq j(i, t)]$, separately by the group of origin $j(i, t) \in \{\text{non-top Firms, Top Firms, Academia}\}$. The next three columns have dep. variable $M_{it} = 1[j(i, t+1) \in \text{Top Firms}]$.

All regressions control for education background (whether bachelor in the U.S., Ph.D. school fixed effect), a cubic polynominal of yrs since PhD as experience (divided by 10), current position types (scientist/engineer/manager), seniority or academic job rank based on job titles on LinkedIn, and firm-year fixed effects. Standard errors are robust and clustered at (Ph.D. school, graduation cohort) level, the unit at which Ph.D.'s are sampled (see Section 4.1-4.2).

Table 6: Effects of Publications & Matched Patents on Additional Mobility Outcomes

Estimates	On New Outputs d_{it}		On Lagged Outputs D_{it}	
	β_{10}	β_{11}	γ_{10}	γ_{11}
Dep. Var: Any Move $j(i, t+1) \neq j(i,t)$				
non-top	0.0351	0.0176	0.0019	0.0180
	(0.0059)	(0.0102)	(0.0036)	(0.0068)
Top	0.0030	-0.0004	-0.0004	0.0034
	(0.0047)	(0.0061)	(0.0032)	(0.0051)
Academia	0.0056	0.0046	0.0078	0.0012
	(0.0023)	(0.0064)	(0.0021)	(0.0044)
Dep. Var: $j(i, t+1) \in$ Top Firms				
non-top	0.0175	0.0129	0.0053	0.0134
	(0.0034)	(0.0036)	(0.0018)	(0.0041)
Top	0.0020	0.0022	0.0013	0.0041
	(0.0038)	(0.0052)	(0.0029)	(0.0044)
Academia	0.0014	0.0021	0.0034	0.0039
	(0.0008)	(0.0022)	(0.0008)	(0.0017)
Dep. Var: $j(i, t+1) \in$ Academia				
non-top	0.0147	0.0093	0.0056	0.0079
	(0.0027)	(0.0045)	(0.0014)	(0.0027)
Top	0.0058	0.0088	0.0015	0.0013
	(0.0019)	(0.0028)	(0.0012)	(0.0021)
Academia	0.0076	-0.0040	-0.0027	0.0025
	(0.0018)	(0.0050)	(0.0017)	(0.0033)
Dep. Var: Move into Higher-Wage Firms				
non-top	0.0339	0.0275	0.0059	0.0178
	(0.0054)	(0.0088)	(0.0031)	(0.0059)
Top	0.0051	0.0093	0.0019	0.0097
	(0.0035)	(0.0050)	(0.0023)	(0.0037)
Academia	0.0026	-0.0008	0.0141	0.0136
	(0.0021)	(0.0050)	(0.0019)	(0.0040)
Dep. Var: Any Promotion				
non-top	0.0315	0.0379	0.0081	0.0231
	(0.0062)	(0.0129)	(0.0035)	(0.0083)
Top	0.0311	0.0157	0.0010	0.0081
	(0.0063)	(0.0070)	(0.0035)	(0.0056)
Academia	0.0435	0.0493	0.0135	0.0253
	(0.0031)	(0.0104)	(0.0025)	(0.0068)

Notes: This table presents estimates of 5.1 for more mobility outcomes. (β_{10}, β_{11}) capture mobility responses to new innovation outputs, without or with a matched patent, respectively. $(\gamma_{10}, \gamma_{11})$ capture mobility responses to lagged innovation outputs, without or with a matched patent, respectively. Higher-wage firms are those with a higher wage for foreign workers (see Appendix B). Any promotion is defined as an ascension to a more senior role, or getting tenured in academia, based on job titles.

Table 7: Effects of Papers & Patents on Job Mobility –Alternative Measure of Info Disclosure

	Move between Firms			Move into Top Firms		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
(Any Pub, Any Patented Pub) at t						
$d_{it}(1,0)$	0.0353	0.0035	0.0089	0.0194	0.0024	0.0031
	(0.0059)	(0.0046)	(0.0020)	(0.0034)	(0.0037)	(0.0007)
$d_{it}(1,1)$	0.0146	0.0004	0.0042	0.0148	0.0026	0.0039
	(0.0104)	(0.0062)	(0.0062)	(0.0064)	(0.0052)	(0.0022)
USPTO Publication at t						
New Patent Revealed	0.0328	0.0043	0.0061	0.0128	0.0001	0.0013
	(0.0113)	(0.0068)	(0.0060)	(0.0064)	(0.0059)	(0.0019)
New Patent Granted	-0.0225	-0.0098	-0.0009	-0.0115	0.0129	-0.0014
	(0.0152)	(0.0091)	(0.0073)	(0.0087)	(0.0077)	(0.0021)
N	222104	62145	121116	222153	62147	121124
Adjusted R2	.1033979	.0156952	.1066973	.0296369	.010779	-.0331524

Notes: This table shows regression estimates of an alternative specification that replaces lagged outputs $D_{it}(\cdot,\cdot)$ by an indicator for any USPTO patent (application) publication at t. It uses the publication dates listed on a patent application whenever available. It also includes an indicator for any new patent granted for comparison. All other controls are the same as in equation 5.1 (see notes under Table 5 for details).

Table 8: Difference in Innovation Outputs in $[t + 1, t + 3]$

	$(1,1)$ vs. $(1,0)$ Stayers $\hat{\delta}_{11,S} - \hat{\delta}_{10,S}$	$(1,1)$ vs. $(1,0)$ Movers $\hat{\delta}_{11,M} - \hat{\delta}_{10,M}$
1. Num. Papers in $[t + 1, t + 3]$		
non-top	0.2486	0.1295
	(0.0609)	(0.1607)
Top	0.2949	0.2552
	(0.0586)	(0.1718)
Academia	0.3748	0.1735
	(0.0558)	(0.1124)
2. Num. Patented Papers in $[t + 1, t + 3]$		
non-top	0.6178	0.2452
	(0.0747)	(0.2191)
Top	0.6534	0.6330
	(0.0836)	(0.2681)
Academia	1.6042	0.8961
	(0.0802)	(0.2889)

Notes: This table shows regression estimates of the future $[t + 1, t + 3]$ productivity difference between $(1,1)$ and $(1,0)$ workers, following equation 5.3. More complete regression results are shown in Table C5.

Parameter	Description	Calibration	ML Estimate
I. Nuisance Parameters - Distributions of Priors			
(μ_H, σ_H)	$ln(\pi_{i1}/(1-\pi_{i1})) \sim \mathcal{N}(\mu_H, \sigma_H)$ if $z_{i1} = 1$		$(-0.66, 0.50)$
(μ_L, σ_L)	$ln(\pi_{i1}/(1-\pi_{i1})) \sim \mathcal{N}(\mu_L, \sigma_L)$ if $z_{i1} = 0$	mixing weights z_{i1} calibrated in Table D1	$(-1.52, 0.20)$
II. Model Parameters			
1. Labor Supply - Preferences for Employers			
b	utility weight on log wage, rel. to idiosyncratic preferences		0.76
ρ_G	$1-$ corr. of ϵ_{itj} for $j \in$ nest $G \in$ {A,P,J,N}		$(0.80, 0.80, 0.60, 0.79)$
$(\eta_{1,G}, \eta_{2,G})$	preference for market G: $\eta_{1,G}\pi + \eta_{2,G}\pi^2$	$\eta_{\cdot,P} = (0,0)$	$\eta_{\cdot,A} = (-0.5, 2.0)$ $\eta_{\cdot,J} = (-0.2, 0.8)$ $\eta_{\cdot,N} = (0.3, -1.4)$
2. Labor Market Dynamics			
$(\lambda_{0,G}, \lambda_{1,G})$	prob. of workers $\in G$ looking for new jobs $\lambda_G(\pi) = \lambda_{0,G} \times (1 + \lambda_{1,G} \times \pi)$, at $t > 1$	$\lambda_{\cdot,P} = (0.38, 0.40)$	$\lambda_{\cdot,A} = (0.07, 0.00)$ $\lambda_{\cdot,J} = (0.10, -0.24)$ $\lambda_{\cdot,N} = (0.09, 1.50)$
Λ_{AJ}	prob. academia is open to workers from industry		0.75
Λ_{JA}	prob. industry is open to workers from academia		0.27
3. Firm Productivity			
f_j	firm j's productivity in routine tasks. 16 employers listed in Section 6.1.1.	first in each nest: $f_0 = 1, f_2 = 0.1, f_4 = 0.15, f_{10} = 1.2$	See Table 10
g_j	firm j's proportional $\uparrow$ in productivity in innovation tasks	first in each nest: $g_0 = 5, g_2 = 3.5, g_4 = 4.0, g_{10} = 2.5$	see Table 10
θ	proportional increase in profits from a high-quality innovation		0.61
(ζ_0, ζ_1)	cost of innovation: $c(\pi, \tau) = \frac{\zeta_0 * (1 + \zeta_1 \times \pi)}{2} \tau^2$	$(3.5, -0.5)$	
4. Worker Productivity			
h	prob. of a H-ability producing a paper ($y = 1$)		0.69
l	prob. of a L-ability producing a paper ($y = 1$)		0.36
$\widetilde{h}_G$	prob. of a H-ability producing a paper with a matched patent ($\widetilde{y} = 1$)		$(0.09, 0.09, 0.48, 0.45)$
$\widetilde{l}_G$	prob. of a L-ability producing a paper with a matched patent ($\widetilde{y} = 1$)		$(0.04, 0.04, 0.14, 0.14)$
5. Others			
β	exponential discount factor of employers	0.90	

Notes: {A,P,J,N} refer to the four nests of employers: Academia - Tenure Track, Postdoc, Top Firms in Industry, non-top Firms in Industry, respectively. There are 4 nuisance parameters and 52 model parameters estimated by maximizing the joint likelihood of job movements and innovation outputs (6.2). See Section 6.1.2 - 6.1.3 for estimation details.

Table 10: Firm Level: Estimated Productivity, Size and Wage Returns

Employer j	Initial Share p_{1j}	Estimated $\widehat{p}_{1j}$	Routine $\hat{f}_j$	Innovation $\hat{g}_j$	Wage Return to Belief π
Nest 1. Tenure Track in Academia					
non-top CS	0.119	0.118	1.000	5.000	0.773
Top 25 CS	0.020	0.022	0.047	9.810	1.082
Nest 2. Postdoc in Academia					
non-top CS	0.073	0.073	0.100	3.500	0.533
Top 25 CS	0.048	0.048	0.040	6.179	1.064
Nest 3. Top Tech in Industry					
IBM	0.039	0.020	0.150	4.000	1.285
Microsoft	0.035	0.024	0.185	3.706	1.221
Amazon	0.007	0.011	0.110	3.086	1.022
Facebook (Meta)	0.003	0.010	0.103	3.214	1.075
Apple	0.004	0.013	0.145	1.413	0.321
Google (Alphabet)	0.043	0.052	0.453	2.859	0.922
Nest 4. non-top in Industry					
SF Bay Area	0.196	0.186	1.200	2.500	0.441
Seattle/LA/Texas	0.142	0.137	0.814	2.409	0.376
NY/Boston/DC/Chicago	0.089	0.095	0.541	2.432	0.367
Other Locations in the US	0.143	0.133	0.784	2.568	0.445
Foreign	0.015	0.029	0.139	2.451	0.348
Missing Location	0.024	0.028	0.134	1.492	0.002

Notes: This table shows actual and predicted share of workers at each employer (size) at $t = 1$, the estimates of firm productivity - $\hat{f}_j$ on routine tasks, $\hat{g}_j$ on innovation tasks, respectively, and the wage return to employer belief π about whether a worker is H-ability. See Section 6.2 for estimation details. The initial allocation of workers across employers, p_{1j} vs. $\hat{p}_{1j}$, is plotted in Figure 7. The relationship between wage returns and an employer's innovation productivity in equilibrium is shown in Appendix Figure D3.

Table 11: Equilibrium Outcomes under Asymmetric vs. Symmetric Information

	Asymmetric Mean (SD)	Symmetric Mean (SD)	Percent Change
Innovation Rates:			
Any Paper	0.094 (0.001)	0.099 (0.001)	4.73% (1.4%)
Any Paper & Patent	0.018 (0.0006)	0.019 (0.001)	6.23% (3.3%)
Talent Revelation:			
$t = 1$	0.182 (0.009)	0.182 (0.009)	0 (.)
$t = 3$	0.439 (0.011)	0.464 (0.011)	5.60% (3.3%)
$t = 5$	0.530 (0.010)	0.554 (0.011)	4.59% (2.6%)
$t = 10$	0.613 (0.011)	0.651 (0.010)	6.19% (2.3%)

Notes: This table shows the rates at which workers produce a paper, or a paper with a matched talent, and the revelation of H-ability workers under asymmetric versus symmetric information. I forward simulate the employment paths and innovation production by workers under each scenario 100 times, holding fixed the initial prior drawn from (6.1) in each round of simulation. See Section 6.3 for details about the counterfactual analysis.

10 Appendix

A. Proofs and Model Extension

A1. Backward Induction

Workers on the Market

Workers who are on the market can choose a new employer as discussed in Section 3.2.1 (see equation 3.7). The choice of an employer is summarized by a static nested logit model. Given a choice set C, workers on the market draw idiosyncratic preferences $\{\epsilon_{itj}\}$ from a GEV distribution (3.2).

Conditional on contracts $\{w_j, \tau_j\}$, define the inclusive value of a nest G of employers as:

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

Therefore, the choice probabilities that enter the labor supply can be written as:

$$p_{j|C} = \underbrace{p_{j|G(j)}}_{\text{choose } j \in G(j)} \times \underbrace{p_{G(j)|C}}_{\text{choose nest } G(j) \in C} \tag{10.1}$$

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

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

Last Period $t = T$

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

$$V_{Tj}\left(\bigcup_{\text{worker } i} I_{iTj}\right) = \underbrace{\sum_{i : j(i,T-1)=j} v_{Tj}^{(1)}(I_{iTj})}_{\text{Incumbent}} + \underbrace{\sum_{i : j(i,T-1)\neq j} v_{Tj}^{(0)}(I_{iTj})}_{\text{Workers Outside}} \tag{10.2}$$

where I_{iTj} represents the information employer j has about worker i at the beginning of T. Employers derive optimal contracts for incumbent versus new workers separately, due to the differences in their labor supply and information about their ability. For an incumbent employee, employer j solves:

$$v_{Tj}^{(1)}(I_{iTj}) = max_{w,\tau} \underbrace{p_j^{(1)}(w; I_{iTj})}_{\text{labor supply}} \times \underbrace{(MP_j(\widetilde{\pi}, \tau) - w)}_{\text{MRPL net wage}} \tag{10.3}$$

$$\text{where } p_j^{(1)}(w, I_{iTj}) = \underbrace{1 - \lambda_{G(j)}(\pi)}_{\text{off market}} + \underbrace{\lambda_{G(j)}(\pi) \times E_C[p_{j|C}(w, w_{(-j)})]}_{\text{on market \& choose j again}}$$

where j's private belief $\widetilde{\pi} = Pr\left(H \,|I_{iTj}\right)$ enters the expected marginal revenue product of a worker (3.1). Employers cannot observe who is on the market before setting the contract. Public belief $\pi = Pr\left(H \,|I_{iT}\right)$ affects the chance of a worker searching for new jobs as well as offers from other employers.

Taking derivatives of the objective function (10.3) over wage w:

$$\frac{\partial p_j^{(1)}(w; I_{iTj})}{\partial w} \times \left(MP_j(\widetilde{\pi}, \tau) - w\right) - p_j^{(1)}(w; I_{iTj}) = 0 \tag{10.4}$$

letting $G = G(j)$, given workers' problem we have,

$$\frac{\partial p_j^{(1)}(w; I_{iTj})}{\partial w} = \lambda_G(\pi) \times \left(\underbrace{\frac{\partial p_{j|G}}{\partial w} \times E_C[p_{G|C}]}_{(a)} + \underbrace{p_{j|G} \times \frac{\partial E_C[p_{G|C}]}{\partial w}}_{(b)} \right)$$

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

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

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

$$\xi_j^{(1)} := \frac{\partial ln(p_j^{(1)}(w; I_{iTj}))}{\partial ln(w)} = \frac{b}{\rho_G} \times E_C \left[\underbrace{\frac{\lambda_G \times p_{j|G} \times p_{G|C}}{p_j^{(1)}}}_{(c)} \times \underbrace{(1 - \rho_G \, p_{j|G} \, p_{G|C} - (1 - \rho_G)p_{j|G})}_{(d)} \right] \tag{10.5}$$

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

Plugging $\xi_j^{(1)}$ into the first order condition (10.4), the optimal wage for an incumbent worker is:

$$w_{iTj}^{(1)} = MP_j(\widetilde{\pi}, \tau_{iTj}^{(1)}) \times \underbrace{\xi_j^{(1)} \times \left(1 + \xi_j^{(1)}\right)^{-1}}_{\text{markdown}} \tag{10.6}$$

Taking the derivative of (10.3) over allocation to innovation tasks, τ,

$$\underbrace{\frac{\partial p_j^{(1)}}{\partial \tau}}_{=0} + p_j^{(1)} \times f_j \left(g_j \, q(\widetilde{\pi}) - 1 - \zeta \tau\right) \geq 0 \tag{10.7}$$

$$\longrightarrow \tau_{iTj}^{(1)} = max\{0, min\{1, \frac{1}{\zeta}\left(g_j \times q(\widetilde{\pi}) - 1\right)\}\}$$

For an outside worker i from $j(i, T-1) \neq j$, employer j has access to information I_{iTj}, which does not indicate if a paper produced during $T-1$ is high-quality. The value function is therefore expected over the quality indicator $\widetilde{y}$. Specifically, employer j solves solves:

$$v_{Tj}^{(0)}(I_{iTj}) = max_{w,\tau} E_{\widetilde{y}} \left[p_j^{(0)}(w; I_{iTj}, \widetilde{y}) \times \underbrace{\left(MP_j(\widetilde{\pi}, \tau) - w\right)}_{\text{MRPL net wage}} \right] \tag{10.8}$$

$$\underbrace{}_{\text{labor supply}}$$

$$\text{where } p_{tj}^{(0)}(w; I_{iTj}, \widetilde{y}) = \lambda_{G_0}(\pi) \times E_C \left[p_{j|C}(w, w_{(-j)})| \widetilde{y}\right]$$

G_0 denotes the original nest worker i is from, $j(i, T-1) \in G_0$, whereas $j \in G$. When the unknown $\widetilde{y} = 1$, the worker's original employer with that information could revise upward the expected marginal revenue product and set a higher wage (10.6). Therefore, the worker is less likely to leave for j, i.e. $p_{tj}^{(0)}(\cdot; 1) < p_{tj}^{(0)}(\cdot; 0)$.

Taking derivatives of (10.8) over wage w:

$$E_{\widetilde{y}} \left[\frac{\partial p_j^{(0)}(w; I_{iTj}, \widetilde{y})}{\partial w} \times \left(MP_j(\widetilde{\pi}, \tau) - w\right) - p_j^{(0)}(w; I_{iTj}, \widetilde{y}) \right] = 0 \tag{10.9}$$

Conditional on the not-yet-revealed $\widetilde{y}$ we have:

$$\frac{\partial p_j^{(0)}(w; I_{iTj}, \widetilde{y})}{\partial w} = \lambda_{G_0}(\pi) \times \left(\underbrace{\frac{\partial p_{j|G}}{\partial w}}_{(e)} \times E_C[p_{G|C}] + p_{j|G} \times \underbrace{\frac{\partial E_C[p_{G|C}]}{\partial w}}_{(f)} \right)$$

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

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

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

$$\xi_j^{(0)}(\widetilde{y}) = \frac{\partial ln(p_j^{(0)}(w; I_{iTj}, \widetilde{y}))}{\partial ln(w)} = \frac{b}{\rho_G} \times E_C \left[\frac{p_{G|C}}{E_C[p_{G|C}]} \times \left(1 - \rho_G \, p_{j|G} \, p_{G|C} - (1 - \rho_G) p_{j|G}\right) \right] \tag{10.10}$$

where each choice probability is conditional on the unknown $\widetilde{y}$. In contrast with the elasticity $\xi_j^{(1)}$ of an incumbent worker (10.5), there is no longer a ratio of the probability of choosing j on the market to that of staying at j, which can be tiny when most incumbent workers do not search for

new jobs. $\lambda_{G_0}(\pi)$, the probability of a worker from nest G_0 getting on the market, does not matter for her elasticity to a new employer j. The other terms in the elasticity are similar, except that for new workers, j is uncertain about $\widetilde{y}$ and considers possible $\epsilon_j^{(0)}(\widetilde{y})$ that are specific to $\widetilde{y}$. New workers are therefore paid higher wages than equally productive incumbent workers, due to their more elastic labor supply. The wages between the two groups would be equal if all incumbent workers search for new jobs.

Multiplying both sides of the FOC (10.9) by $\frac{w}{E_{\widetilde{y}}[p_j^{(0)}(\widetilde{y})]}$ and plugging in $\epsilon_j^{(0)}(\widetilde{y})$'s, we have :

$$
E_{\widetilde{y}}\left[\frac{p_j^{(0)}(\widetilde{y})}{E_{\widetilde{y}}[p_j^{(0)}(\widetilde{y})]} \times \left(\xi_j^{(0)}(\widetilde{y}) \times (MP_j(\widetilde{\pi}, \tau_j^{(0)}) - w) - w \right) \right] = 0
$$

note by Bayes Rule, the weight on $\widetilde{y} = 1$: $Pr(\widetilde{y} = 1|\pi) \times \dfrac{p_j^{(0)}(\widetilde{y})}{E_{\widetilde{y}}[p_j^{(0)}(\widetilde{y})]} = Pr(\widetilde{y} = 1|\pi, \text{worker chooses } j)$

Therefore, the optimal wage for a new worker is:

$$
\longrightarrow w_{iTj}^{(0)} = E_{\widetilde{y}}[MP_j(\widetilde{\pi}, \tau_{iTj}^{(0)}) \times \xi_j^{(0)}(\widetilde{y}) \mid \text{choose j}] \times \left(1 + E_{\widetilde{y}}[\xi_j^{(0)}(\widetilde{y}) \mid \text{choose j}] \right)^{-1} \tag{10.11}
$$

the expectations in which are over $\widetilde{y}$ conditional on a worker choosing j at period $t = T$.

Taking the derivative of (10.8) over task allocation τ,

$$
E_{\widetilde{y}}[p_j^{(0)}(\widetilde{y}) \times f_j \times (g_j\, q(\widetilde{\pi}) - 1 - \zeta\tau)] \geq 0 \tag{10.12}
$$

$$
\longrightarrow E_{\widetilde{y}}[g_j\, q(\widetilde{\pi}) - 1 - \zeta\tau \mid \text{choose j}] \geq 0
$$

$$
\longrightarrow \tau_{iTj}^{(0)} = max\{0, min\{1, \frac{1}{\zeta}(g_j\, E_{\widetilde{y}}[q(\widetilde{\pi}) \mid \text{enter j}] - 1)\}\}
$$

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

Employers now take into the expected continuation value from stayers at $(t + 1)$. Letting $I(y\widetilde{y}) \in \{I(00), I(10)\, I(11)\}$ denote potential information set tomorrow conditional on innovation outputs $(y, \widetilde{y})$ produced at t, employer j solves the following for incumbent workers:

Incumbent $j = j(i, t-1)$: $v_{tj}^{(1)}(I_{itj}) = max_{w, \tau}\, \underbrace{p_j^{(1)}(w;\, I_{itj})}_{\text{labor supply}} \times \underbrace{\left(MP_j(\widetilde{\pi},\, \tau) + \beta \times E[v_{(t+1)j}^{(1)}(I) \mid I_{itj}, \tau] - w \right)}_{\text{MRPL at t and discounted continuation value, net wage}}$

$$\tag{10.13}$$

in which the expected continuation value is:

$$
E\left[v^{(1)}_{(t+1)j}(I)\mid I_{itj},\tau\right] = E_{(y,\widetilde{y})}\left[E[v^{(1)}_{(t+1)j}(I)\mid(y,\widetilde{y}),I_{itj},\tau]\right] \tag{10.14}
$$

$$
= \underbrace{\tau \times q_{11}(\widetilde{\pi}) \times v^{(1)}_{(t+1)j}(I(11))}_{\text{output at t : (1,1)}} + \underbrace{\tau \times q_{10}(\widetilde{\pi}) \times v^{(1)}_{(t+1)j}(I(10))}_{\text{output at t : (1,0)}}
$$

$$
+ \underbrace{(1 - \tau \times (q_{11}(\widetilde{\pi}) + q_{10}(\widetilde{\pi}))) \times v^{(1)}_{(t+1)j}(I(00))}_{\text{output at t : (0,0)}}
$$

$$
\text{where } q_{11}(\widetilde{\pi}) = \widetilde{\pi} \times h \times \widetilde{h} + (1 - \widetilde{\pi}) \times l \times \widetilde{l}
$$

$$
\text{and } q_{10}(\widetilde{\pi}) = \widetilde{\pi} \times h \times (1 - \widetilde{h}) + (1 - \widetilde{\pi}) \times l \times (1 - \widetilde{l})
$$

The optimal wages at $t < T$, as shown in (3.12) and repeated below, can be derived the same way as wages at $t = T$.

$$
w^{(1)}_{itj} = \underbrace{\left(MP_j(\widetilde{\pi},\tau^{(1)}_{itj}) + \beta \times E[v^{(1)}_{(t+1)j}(I)\mid I_{itj},\tau^{(1)}_{itj}]\right)}_{\text{MRPL plus continuation value}} \times \underbrace{\xi^{(1)}_j \times \left(1 + \xi^{(1)}_j\right)^{-1}}_{\text{markdown}}
$$

The firm-specific labor supply elasticity of an incumbent worker vs. a new worker can be written the same as equations (10.5) (10.10), respectively. The ξ's are functions of wages posted by potential employers, and since the worker's problem (if on the market) remains the same each period, the elasticity function is also time-invariant.

The difference in wages at $t < T$ from $t = T$ ones is that employers also share some of the expected continuation value with the worker (marked down by the inverse of labor supply elasticity). In other wages, the dynamic monopsonistic wages in this framework are front-loaded. Once a worker has entered the firm, wages for incumbent employees are lower. The gap between an incumbent and equally productive new worker may be interpreted as a signing bonus or stock options contracted upon entry.

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

$$
f_j\left(g_j\,q(\widetilde{\pi}) - 1 - \zeta\tau\right) + \beta \times \frac{\partial E[v^{(1)}_{(t+1)j}(I)\mid I_{itj},\tau]}{\partial\tau} \geq 0 \tag{10.15}
$$

$$
\text{where } \frac{\partial E[v^{(1)}_{(t+1)j}(I)\mid I_{itj},\tau]}{\partial\tau} = q_{10}(\widetilde{\pi}) \times \left(v^{(1)}_{(t+1)j}(I(10)) - v^{(1)}_{(t+1)j}(I(00))\right) + q_{11}(\widetilde{\pi}) \times \left(v^{(1)}_{(t+1)j}(I(11)) - v^{(1)}_{(t+1)j}(I(00))\right)
$$

Therefore, we reach the optimal allocation to innovation tasks in equation (3.13), and repeated below:

$$
\tau^{(1)}_{itj} = max\{0, min\{1, \frac{1}{\zeta}(\underbrace{g_j \times q(\widetilde{\pi}) - 1}_{\text{returns to innov today}} + \beta \times \underbrace{\partial E[v^{(1)}_{(t+1)j}(I)\mid I_{itj},\tau]/\partial\tau}_{\text{change in continuation value}})\}\}
$$

The optimal contracts for a new worker maximize (3.14). The derivation is similar to that of $t = T$. When considering continuation value, employers have to consider what $\widetilde{y}$ at the previous

employer, which is unknown but will be revealed by the beginning of $(t + 1)$.

In summary, we have derived the optimal wages as expressed in (3.12,3.16), and the optimal task allocations in (3.13,3.17). In equilibrium, employers set wages and allocate workers to innovation tasks, conditional on information about workers and taking as given the wages set by other employers. The expected labor supply from incumbent and new workers is determined by the wages set by potential employers.

First Period $t = 1$

Since firms begin with no employee, we do not need separate value functions for incumbent vs. new workers. Firms post contracts based on common initial information $I_{i1j} \equiv I_{i1}$ about a worker that yields belief $\pi = Pr(H|I_{i1})$:

$$V_{1j}\left(\bigcup I_{i1}\right) = \sum_i v_{1j}(I_{i1}) \tag{10.16}$$

$$v_{1j}(I_{i1}) = max_{w,\tau} \underbrace{p_j(w; I_{i1})}_{\text{labor supply}} \times \left(\underbrace{MP_{1j}(\pi, \tau)}_{\text{MRPL at } t=1} + \underbrace{\beta \times E[v_{2j}^{(1)}(I)|I_{i1}, \tau]}_{\text{continuation value}} - w \right)$$

All workers are on the market at $t = 1$ and can choose from any employer. The FOC for initial wage can be simplified from equations (10.4, 10.9):

$$\frac{\partial p_j(w; I_{i1})}{\partial w} \times \left(MP_j(\pi, \tau) - w\right) - p_j(w; I_{i1}) = 0$$
$$\text{where } p_j(w; I_{i1}) = p_{j|G} \times p_G$$

Define the initial elasticity to firm $j \in G$ as:

$$\epsilon_{1j}(I_{i1}, w) = \frac{b}{\rho_G} \times \left(1 - (1 - \rho_G)p_{j|G} - \rho_G p_j\right) \tag{10.17}$$

The optimal contract can then be written as:

$$w_{i1j} = \left(MP_j(\pi, \tau_{i1j}) + \beta E[v_{2j}^{(1)}(I)| I_{i1}, \tau_{i1j}]\right) \times \epsilon_{1j} \times \left(1 + \epsilon_{1j}\right)^{-1} \tag{10.18}$$

$$\tau_{i1j} = max\{0, min\{1, \frac{1}{\zeta}\left(g_j \times q(\pi) - 1 + \beta \times \frac{\partial E[v_{2j}^{(1)}(I)|I_{i1}, \tau]}{\partial \tau}\right)\}\}$$

where wage markdown equals the inverse of labor supply elasticity in (10.17), and the continuation value changes in τ as in (10.15).

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

A2. Proof of Propositions 1 and 2

Proof of Proposition 1 - Unique Equilibrium under Monopsonistic Competition

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

$$
w^*_{itj} = \begin{cases} w^{(1)}_{tj}(I_{it} \cup \{\widetilde{y}_{i(t-1)}\}) & j = j(i, t-1), \text{ as in equations (3.12, 10.6)} \\ w^{(0)}_{tj}(I_{it}) & j \neq j(i, t-1), \text{ as in equations (3.16, 10.11)} \end{cases}
$$

In equilibrium, the probability of a worker on market choosing employer j, as expressed in (10.1), is determined by the wages set by all potential employers:

$$
p^*_{ij|C} = \underbrace{\frac{exp\left(\eta_{G(j)}(I_{it}) + \rho_{G(j)}W^*_{iG(j)}\right)}{\sum_{G \in C} exp\left(\eta_G(I_{it}) + \rho_G W^*_{iG}\right)}}_{\text{choose nest } G(j) \text{ in choice set } C} \times \underbrace{\frac{exp\left(b/\rho_{G(j)} \, ln(w^*_{itj})\right)}{exp(W^*_{iG(j)})}}_{\text{choose } j \text{ within nest } G(j)} \tag{10.19}
$$

where the inclusive value for nest G equals $W^*_{iG} := ln\left(\sum_{j \in G} exp(b/\rho_G ln(w^*_{itj}))\right)$. The equilibrium allocation of workers across firms can be viewed as a fixed point of the function $p \circ w$:

$$
p(w(p)) = p
$$

To show that the equilibrium allocation is unique (log wages are unique up to a constant), I prove $M : \mathbf{R}^K \to \mathbf{R}^K$ is a contraction mapping with modulus less than 1:[66]

$$
\forall t \, \forall j > 1 : \; f(\boldsymbol{\omega}_{tj}) = \boldsymbol{\omega}_{tj} + p_{tj} - p_{tj}(\boldsymbol{\omega}_t) \tag{10.20}
$$

$$
\omega_{tj} := \frac{b}{\rho_{G(j)}} ln(\frac{w_{tj}}{w_{1j}}) \tag{10.21}
$$

where ω_{tj} are log wages multiplied by $\frac{b}{\rho} \in (0, \infty)$, relative to that of $j = 1$, and $p_{tj}(\cdot)$ represents the labor supply given wages, which are different for incumbent (3.11) and new workers (3.15).

Following Berry, Levinsohn, and Pakes (1995; henceforth BLP), I show that M satisfies the sufficient conditions for a contraction that are laid out in Theorem 1 of BLP. I focus on the proof for incumbent workers $\in j$, with labor supply (3.11) at $t > 1$. The proof for new workers is similar.

---

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

Given any positive wages, the derivatives satisfy:

$$\frac{\partial M_{tj}}{\partial \omega_{tj}} = 1 - \frac{1}{p_{tj}^{(1)}(\omega_t)}\frac{\partial p_{tj}^{(1)}}{\partial \omega_{tj}} = 1 - \underbrace{\frac{\xi_{tj}^{(1)}}{b/\rho_{G(j)}}}_{\text{see (10.5)}} \geq 0 \tag{10.22}$$

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

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

$$q \in G(j): \frac{\partial p_{tj}^{(1)}}{\partial \omega_{tq}} = \lambda_{G(j)} \times \left(p_{j|G(j)} \times p_{q|G(j)}\right) \times E_C[-p_{G(j)|C} + \rho_{G(j)} \times p_{G(j)|C}(1 - p_{G|C})] \tag{10.23}$$

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

The sum of the derivatives in (10.22) for each $j > 1$, at each period $t > 1$:

$$\sum_{q>1}\frac{\partial M_{tj}}{\partial \omega_{tq}} = 1 + \frac{\lambda_{G(j)} \times p_{j|G(j)}}{p_{tj}^{(1)}} \times \left(E_C[-p_{G(j)|C} \times (1 - \rho_{G(j)}p_{j|C} - (1 - \rho_{G(j)})p_{j|G(j)})]\right. \tag{10.24}$$

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

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

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

$$+ \rho_{G(j)} \times \underbrace{\sum_{q>1} p_{q|C} \times p_{G(j)|C} - p_{G(j)|C}}_{<1}]$$

which satisfies

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

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

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

unique fixed point such that

$$\forall t \,\forall j > 1: \boldsymbol{\omega}_{tj}^* = \boldsymbol{\omega}_{tj}^* + p_{tj} - p_{tj}(\boldsymbol{\omega}_t^*) \tag{10.25}$$

The fixed point $\boldsymbol{\omega}^*$ can be translated to equilibrium wages that are unique up to (nonzero) scaling factor. The equilibrium allocation of workers across firms, as in (10.19) is unique.

$\square$

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

Given $\frac{b}{\rho} \to \infty$ and $\forall G: \lambda_G \equiv 1$, the labor supply elasticity of incumbent and new workers, as expressed in (10.5) and (10.10) both go to infinity. The labor market is perfectly competitive given that the labor supply of every worker is perfectly elastic w.r.t wages.

Plugging $\xi^{(1)}$ into the wage for incumbents at $t = T$, we have $w_{iTj}^{(1)}(\widetilde{\pi}) = MP_j(\widetilde{\pi}, \tau_{iTj}^{(1)})$. That is, incumbent workers with belief $\widetilde{\pi}$ are paid the full marginal revenue product of labor. There is thus no dynamic rent for employers at $(T-1)$. The wage in intermediary periods, as shown in (3.12), also equals to a worker's MRPL without leaving any rent to an employer.

Information is assumed to be symmetric between employers. The expectation over $\widetilde{y}$, which indicates the quality of a paper (whether it has a matched patent), can be removed from the wages for new workers as in (3.16). Therefore we have,

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

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

Since the continuation value equals zero at all employers, allocating workers to innovation tasks also becomes a static decision. The solutions in (3.17,3.13,10.12,10.7,10.18) can be simplified to:

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

The costs of innovation tasks are fully deducted from workers' wages (see 3.1). That is, workers are bearing all costs of innovation. They are not credit constrained as they earn a positive wage from routine tasks (under Assumption 3 that $\forall j: M_j > 0$). The choices of innovation tasks would be first best in each period, just like the choice of general skill training made by workers who are not credit constrained in Becker (1964).

$\square$

A3. Model Predictions

Derivation of Prediction 1: Mobility in Response to Public Information

Given information I, denote by $\pi_1 = Pr(H|I \cup \{1\})$ the public belief when a worker has any innovation, and by $\pi_0 = Pr(H|I \cup \{0\})$ the belief otherwise. Assumption 3 implies

$$\pi_1 > \pi_0$$

a) According to (3.11), a worker who produces a public innovation stays at the incumbent employer j with probability

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

78

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

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

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

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

$$+ \lambda_{G(j)}(\pi_1) \times \underbrace{\left(p_{j|G(j)}(\pi_1) \times E_C[p_{G|C}(\pi_1)] - p_{j|G(j)}(\pi_0) \times E_C[p_{G|C}(\pi_0)]\right)}_{\text{choose j again if on market}}$$

Under Assumption 2, $\pi_1 > \pi_0 \rightarrow \lambda(\pi_1) \geq \lambda(\pi_0)$. Unless workers with belief π_1 are much more likely to choose the incumbent j again out of all potential employers, the difference above is negative.[67] It implies workers with a new signal are more likely to leave their incumbent employers than similar coworkers without a signal.

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

$$p_{j'}(\pi_1) = \frac{exp(b \times ln(w_{j'}(\pi_1))}{exp(\Omega(\pi_1))}, \quad p_{j'}(\pi_0) = \frac{exp(b \times ln(w_{j'}(\pi_0))}{exp(\Omega(\pi_0))}$$

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

Under Assumption 3 and the optimal solutions shown in (3.12, 3.16, 10.6, 10.11), wages are nondecreasing in belief π, resulting in $\Omega(\pi_1) \geq \Omega(\pi_0)$. Moreover, the wage increase is larger at more productive firms (higher $g_{j'}$) that can allocate more innovation tasks to π_1 than other firms on average. In summary, workers with π_1 are more likely to move into j' if the following conditions hold:

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

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

The positive assortative matching affects marginal workers who would not have spent as much time on innovation task without the positive signal. If π_1, π_0 are significantly high, the worker might be able to spend 100% of time on innovation at any firm, and there is no sorting as in a

---

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

standard AKM framework.[68]

Derivation of Prediction 2: Mobility under Asymmetric Information

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

a) Denote by $\widetilde{\pi}_{11}$ the private belief about worker 1, and $\widetilde{\pi}_{10}$ the private belief about worker 2. Based on the labor supply in (3.11, 3.15), at the beginning of t the difference in the probability a worker stays with the incumbent employer j :

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

(10.29)

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

b) Given assumptions on the information structure (3.5, 3.6), $\widetilde{y}_{1(t-1)} > \widetilde{y}_{2(t-1)}$ are revealed by $(t + 1)$. As the market receives more positive signals about worker 1 than 2, Prediction 1 applies and we have the $(1, 1)$ worker more likely to move to a new firm and more productive one than the $(1, 0)$ worker.

Derivation of Prediction 3: Productivity of Movers vs. Stayers

Consider workers with a common public belief but different signals $(1, 1)$ vs. $(1, 0)$ that have not been revealed fully to the outside market.

a) Workers with outputs $(1, 1)$ are more likely to have H-ability than those with $(1, 0)$, under Assumption 3. No matter if they stay with the incumbent employer or not, per unit of time on innovation task $(1, 1)$ are more likely to produce a paper or a high-quality paper with a matched patent application.

b) Among stayers, an incumbent employer can distinguish between $(1, 1)$ and $(1, 0)$. Given private belief $\widetilde{\pi}(11) > \widetilde{\pi}(10)$, we have $\tau_j(\widetilde{\pi}(11)) \geq \tau_j(\widetilde{\pi}(10))$ according to the optimal task allocations in (3.13, 10.7). In contrast, outside employers would assign an equal amount of time to innovation for $(1, 1)$ and $(1, 0)$ workers with the same public belief. The productivity difference between stayers is thus at least as large as the difference between movers.

---

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

$$\forall \pi : ln(w_j(\pi)) = \alpha(\pi) + \phi_j \tag{10.26}$$

$$\rightarrow \frac{\Omega(\pi_1)}{\Omega(\pi_0)} = exp(b(\alpha(\pi_1) - \alpha(\pi_0))) \times 1 \tag{10.27}$$

$$ln\left(\frac{p_{j'}(\pi_1)}{p_{j'}(\pi_0)}\right) = b\ (\alpha(\pi_1) - \alpha(\pi_0)) - b\ (\alpha(\pi_1) - \alpha(\pi_0)) = 0 \tag{10.28}$$

A4. Model Extension - Forward-looking Workers

So far we have assumed workers consider the utility from wage only, which equals the net present value of a worker-firm match, marked down by the inverse of labor supply elasticity (see 10.5, 10.10). In a more general framework, workers can be forward-looking and take into account their option value in the labor market next period if they enter a firm now. Conditional on wages today, working for a more innovative firm would be more appealing to a high-ability individual who can improve the future market belief about her by producing more innovation today.

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

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

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

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

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

$$
\begin{aligned}
u_{itj}(w, \tau; \pi) = & \ b \times ln(w) + \beta_i \times E_{(y, \widetilde{y})}[\Omega_{i(t+1)}(\pi(y, \widetilde{y}) \mid \tau] + \epsilon_{ijt} \\
= & \ b \times ln(w) + \beta_i \times \tau \times (\pi \widetilde{h} h + (1-\pi)\widetilde{l}l) \times \Omega_{i(t+1)}(\pi(1,1)) \\
& + \beta_i \times \tau \times (\pi(1-\widetilde{h})h + (1-\pi)(1-\widetilde{l})l) \times \Omega_{i(t+1)}(\pi(1,0)) \\
& + \beta_i \times (1 - \tau \times (\pi h + (1-\pi)l) \times \Omega_{i(t+1)}(\pi(0,0)) + \epsilon_{ijt}
\end{aligned}
\tag{10.30}
$$

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

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

$$
\begin{aligned}
u_{itj}(w, \tau; \pi) = & \ b \times ln(w) + \gamma(\pi) \times \tau + \epsilon_{ijt} \\
\text{where } \gamma(\pi) = & \beta_i \times (\pi \widetilde{h} h + (1-\pi)\widetilde{l}l) \times \big(\Omega_{i(t+1)}(\pi(1,1)) - \Omega_{i(t+1)}(\pi(0,0))\big) \\
& + \beta_i(\pi(1-\widetilde{h})h + (1-\pi)(1-\widetilde{l})l) \times \big(\Omega_{i(t+1)}(\pi(1,0)) - \Omega_{i(t+1)}(\pi(0,0))\big)
\end{aligned}
\tag{10.31}
$$

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

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

$$
p_j(w, \tau; \pi) = \frac{exp(b \times ln(w) + \gamma(\pi) \times \tau)}{\sum_{j'} exp(b \times ln(w_{j'}) + \gamma(\pi) \times \tau_{j'})}
\tag{10.32}
$$

And the optimal task allocation chosen by j solves:

$$
\frac{\gamma(\pi)}{u'(w)} + \frac{\partial}{\partial \tau}(MP_j(\pi) - w + \beta E[v_{i(t+1)j}|\tau, \pi]) = 0
\tag{10.33}
$$

where the first part is a ratio of the marginal utility of τ vs. wage w, and the second part is the marginal return to spending more time on innovation as in the benchmark model. The benchmark model assumes $\beta_i = 0$, which implies $\gamma(\pi) \equiv 0$, and we are back to the first-order conditions shown in (10.7), for example.

When $\gamma(\pi) > 0$, a worker prefers to spend more time on innovation as it improves her option value in the labor market next period. Equation (10.31) shows γ is nondecreasing in π, which suggests workers who are more likely to be high-ability further sort themselves into firms allocating more innovation tasks.

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

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

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

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

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

B. Data

B1. Publications Data

The main data source of research papers is Scopus, an abstract and citation databases of peer-reviewed literature launched by Elsevier in 2004. For each conference/journal × year, a query is submitted via Scopus Search API, and it returns a list of papers with information such asauthor(s), title, abstract, ISSN, DOI, number of citations, volume, issue, and publication date.

Scopus also provides affiliations IDs at paper × author level. Another query is submitted for each affiliation ID via the Affiliation Search API, and returns the corresponding institution's name and location. To maximize matching with an author's employment history, I used the same script that cleans the names of employers on LinkedIn profiles to harmonize the affiliation names from Scopus. I consider a paper by author i affiliated to j as her on-the-job research if:

1. j can be matched with an employer of i on her LinkedIn profile;

2. Author i is employed by j at the time of publication.

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

To evaluate paper quality, I collected citations from Scopus, which covers both journal articles and conference papers. Citations from other conference papers are particularly important in computer science. Some scientometric studies suggest Scopus has better coverage of conference proceedings when compared to Web of Science (e.g., Harzing 2019, Pranckute 2021). **FLGA** I selected a random sample of papers and looked at the total citation counts on both Scopus and Web of Science (WoS). Indeed Scopus gives a more comprehensive list of citations than WoS.

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

B2. Match between Papers and Patent Applications

I collected patent data from the Patent Examination Research Dataset (PatEx), which contains publicly viewable patent applications from the Public Patent Application Information Retrieval system (Public PAIR) until the end of 2021.[69] For each patent application, I collected the names of inventors, and related parent/child application within a family, dates of the earliest filing, publication of the application, and grant date if a patent is eventually granted. I then merged patent

---

[69]PatEx 2021 "contains detailed information on more than 12.5 million publicly-viewable provisional and non-provisional patent applications to the USPTO and over 1 million Patent Cooperation Treaty (PCT) applications. It is based on data that OCE downloaded from the Patent Examination Data System (PEDS) in June, 2022. The PEDS data are sourced from Public PAIR." https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-examination-research-dataset-public-pair

applications with USPTO's Patent Assignment Dataset to obtain the assignee of an application, which are typically the employer(s) of inventors.[70]

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

1. For each (paper, author) pair in year t, I looked for all (patent app, inventor) with the inventor = author that are initially filed between years $[t-3, t+3]$. Considering the number of authors/inventors matched at the paper/patent level, I drop matches if:

   - Less than half of the inventors on a patent application are matched, and less than half of the authors on a paper are matched.

   - The number of inventors on a potential matched patent is $< 1/3$ or > 3 the number of authors on the paper.

2. Merge the matched (paper, patent, author/inventor) from (1) with author affiliations from Scopus at (paper, author) level, and with assignees at (patent, assignee) level.

   - Keep (paper, author/inventor, patent) matches if the author's affiliation is matched with one of the patent assignees.

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

To further enhance match quality, I compare the titles and abstracts of papers from Scopus, with titles and abstracts for potentially matched patent applications, which are extracted from Google Patents Public Datasets via BigQuery. I used OpenAI's Ada V2 text embedding model to create numerical representations of paper or patent abstracts.[71] Each embedding is a vector of dimension 1,536. The more similar a patent abstract to a paper's, the smaller the distance between their vector embeddings. This measure of paper-patent similarity is available for 85% of the potential matches.

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

1. # matched authors, # matched inventors on a patent in descending order

2. at least one author affiliation can be matched with patent assignee

3. prefers patent application filed in $[t-1, t]$ (t: yr of paper publication)

4. distance between text embeddings, in ascending order

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

---

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

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

Figure B1: Job Postings for Research Scientists

(a) Amazon Science

BASIC QUALIFICATIONS

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

PREFERRED QUALIFICATIONS

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

(b) Google Research

Minimum qualifications:

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

Preferred qualifications:

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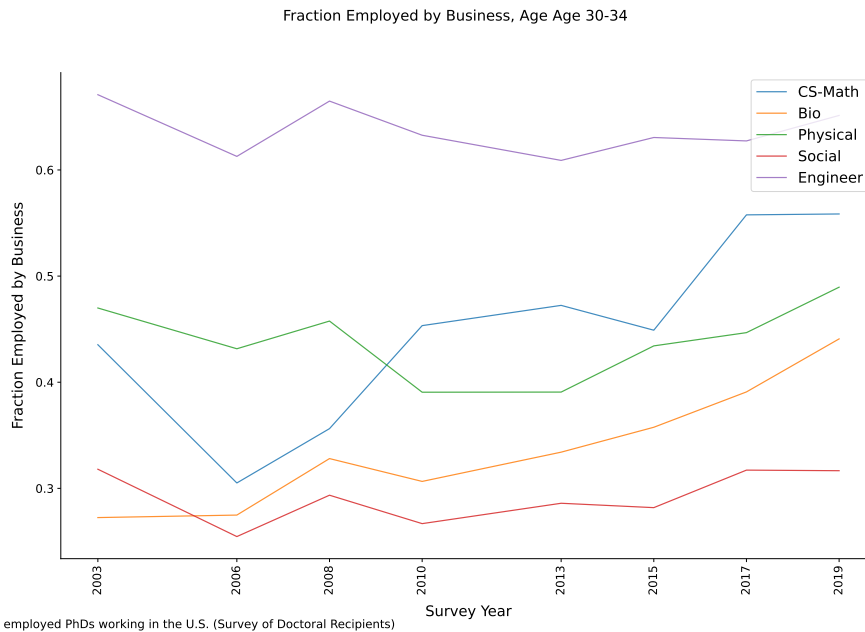- Ability to design and execute on research agenda.

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

Figure B2: CS PhDs

(a) New PhDs (Survey of Earned Doctorates)
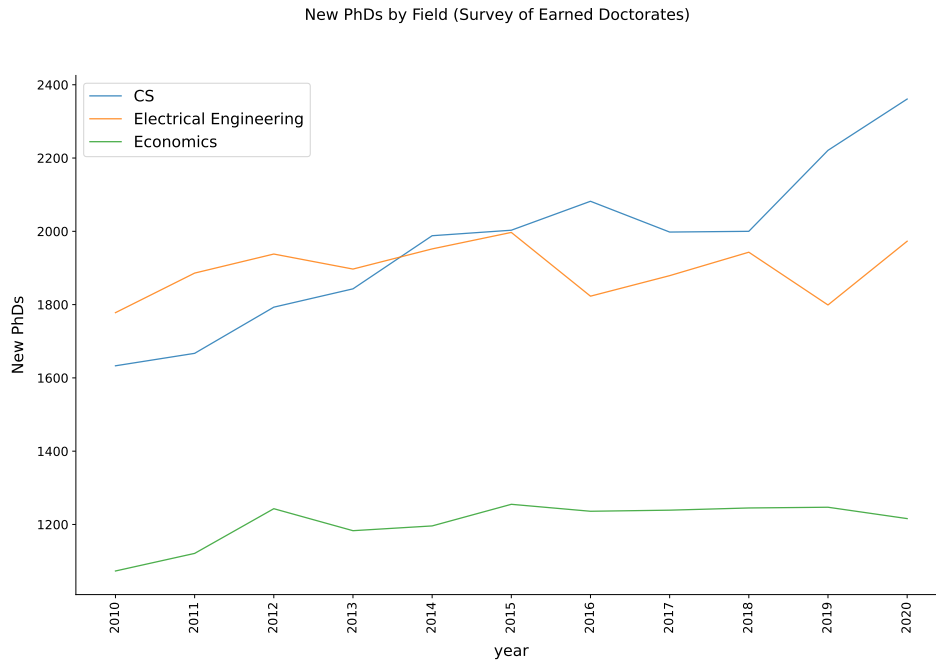
New PhDs by Field (Survey of Earned Doctorates)

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

Fraction Employed by Business, Age Age 30-34

Restrict to employed PhDs working in the U.S. (Survey of Doctoral Recipients)

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

Figure B3: Conference Proceedings

Trend in Papers that have a CS Ph.D. Author
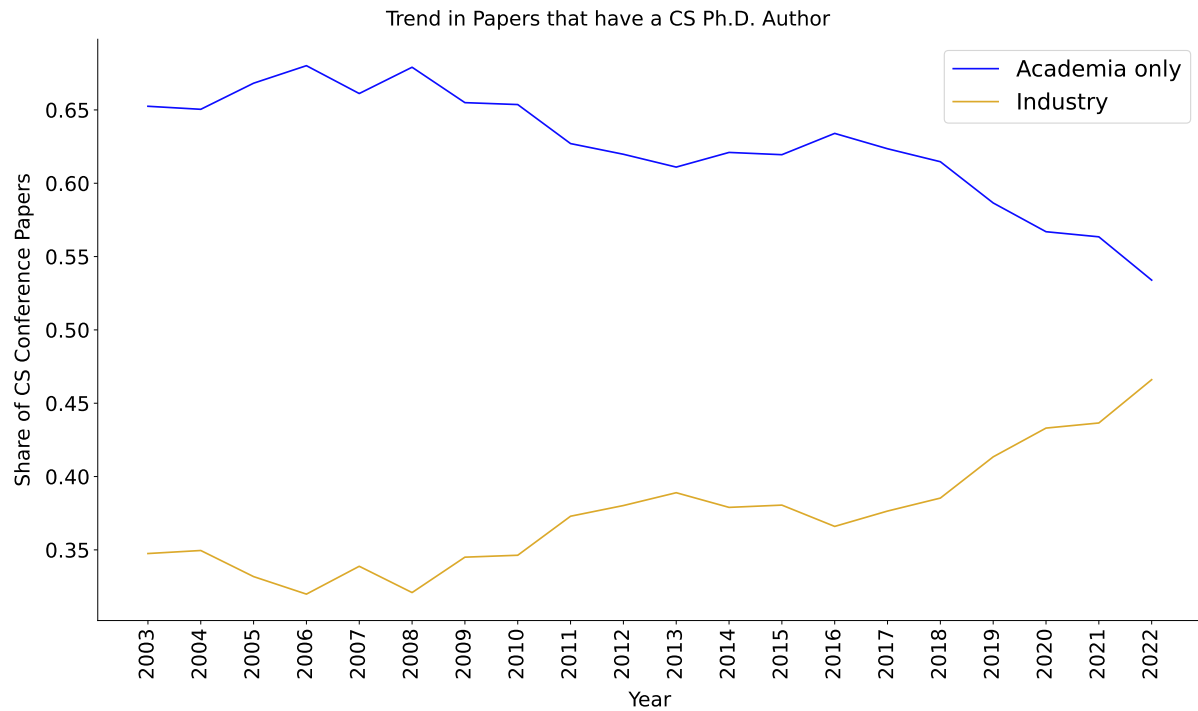
Notes: This figure shows the share of CS conference papers that are purely from academia, versus with at least one author from industry, in each year between 2003 and 2022. I restrict to papers that include a CS Ph.D. author and that have nonmissing author affiliation records. Collaborations between industry and academic researchers are counted in "Industry" (yellow line). The trends are very similar if I further restrict to papers where authors are not affiliated with their Ph.D. institutions, which exclude papers done during grad school or collaborations with former classmates or advisors.

Figure B4: 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 B5: Patents on LinkedIn
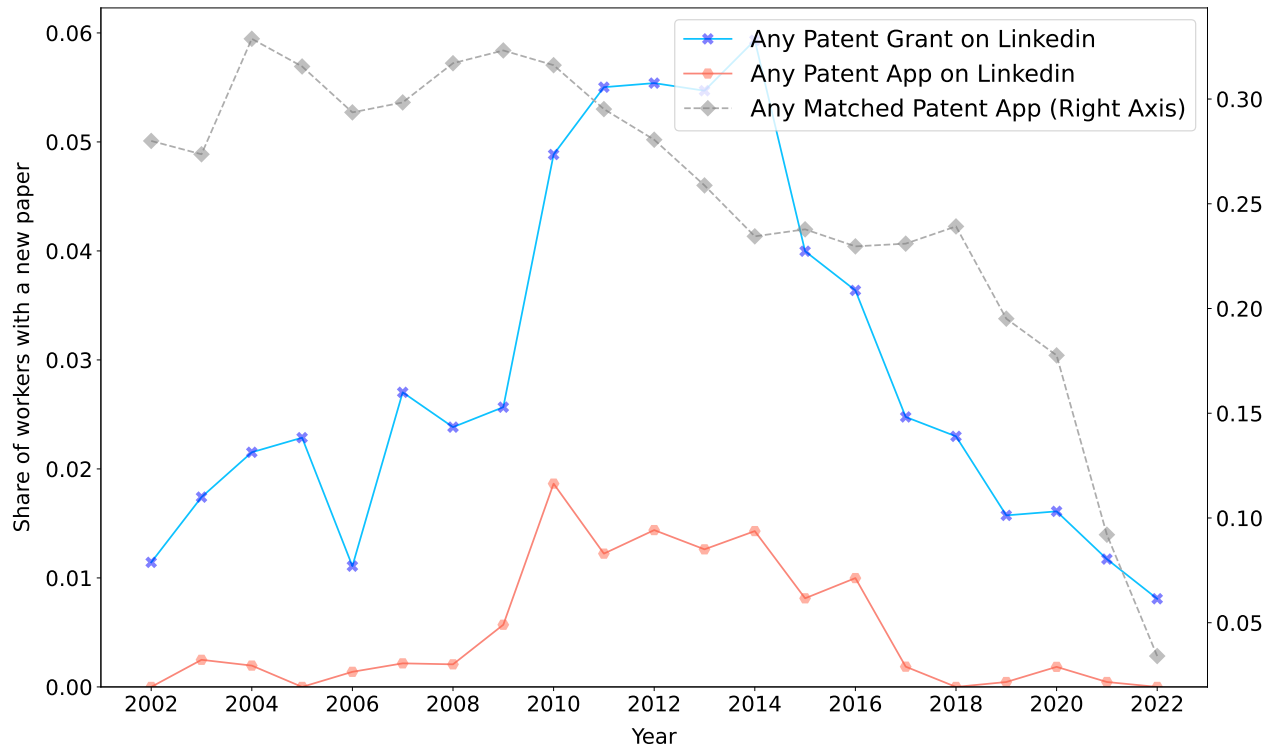
Note 1: This figure shows the time series of the share of workers who listed a patent on LinkedIn by year, conditional on having a new paper that year. The patents section on a LinkedIn profile may include either a patent grant or application, and provides the grant and/or filing date(s). The blue line (left axis) shows the share of workers who have a new paper in a given year (based on publication records) and list a granted patent the same year on LinkedIn. The red line (left axis) shows the share of workers who have a new paper and list a patent application the same year. The gray line (right axis) shows the share of workers who also have a patent application matched to a new paper, for comparison.

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

Figure B6: CS PhDs

(a) Patent Applications

Trend in Patent Applications by Employers in Sample
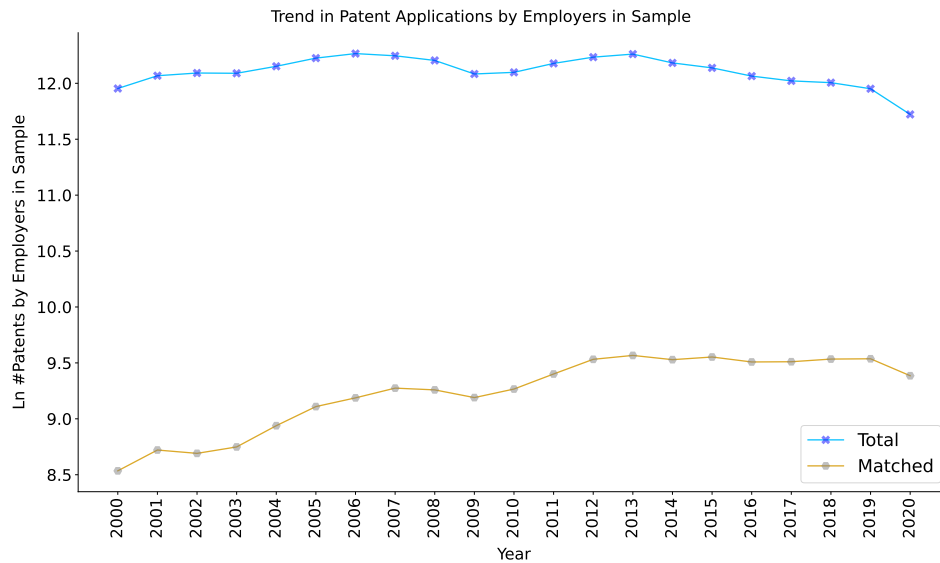
(b) Share of Patents with a Matched CS Paper
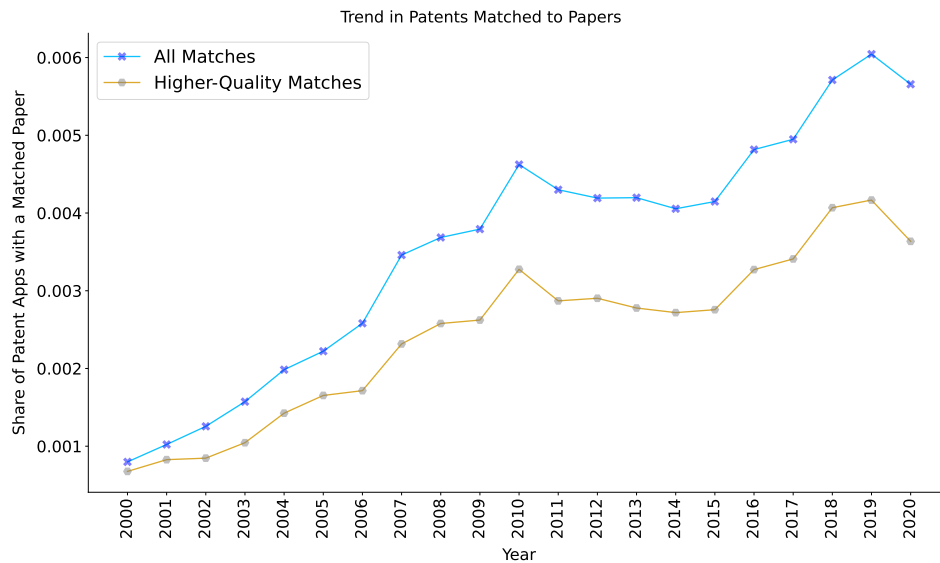
Trend in Patents Matched to Papers

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

Table B1: Number of Profiles by Year

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

Table B2: Number of Profiles by School (ProQuest vs. School-specific Dissertation Database or Websites)

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

Table B3: patent laws - Title 35, United States Code

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

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

C. Additional Empirical Tests for Learning (Section 5)

Table C0: Job Mobility on Publications & Matched Patents (Baseline Specification 5.1)

	Move between Firms			Move into Top Firms		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
(Any Pub, Any Patented Pub) at t						
$d_{it}(1,0)$	0.0351	0.0030	0.0056	0.0175	0.0020	0.0014
	(0.0059)	(0.0047)	(0.0023)	(0.0034)	(0.0038)	(0.0008)
$d_{it}(1,1)$	0.0176	-0.0004	0.0046	0.0129	0.0022	0.0021
	(0.0102)	(0.0061)	(0.0064)	(0.0061)	(0.0052)	(0.0022)
(Any Pub, Any Patented Pub) between $[t-3, t-1]$						
$D_{it}(1,0)$	0.0019	-0.0004	0.0078	0.0053	0.0013	0.0034
	(0.0036)	(0.0032)	(0.0021)	(0.0018)	(0.0029)	(0.0008)
$D_{it}(1,1)$	0.0180	0.0034	0.0012	0.0134	0.0041	0.0039
	(0.0068)	(0.0051)	(0.0044)	(0.0041)	(0.0044)	(0.0017)
Controls W_{it}						
Experience = Yrs since PhD /10	-0.0017	0.0277	-0.0340	-0.0022	-0.0228	-0.0015
	(0.0060)	(0.0084)	(0.0086)	(0.0026)	(0.0075)	(0.0023)
Square of Experience	-0.0197	-0.0271	0.0023	-0.0054	0.0206	-0.0008
	(0.0045)	(0.0070)	(0.0049)	(0.0018)	(0.0064)	(0.0013)
Cube of Experience	0.0045	0.0052	0.0011	0.0011	-0.0044	0.0002
	(0.0009)	(0.0016)	(0.0009)	(0.0004)	(0.0015)	(0.0002)
Yrs in Academia since PhD /10	-0.0002	0.0039	-0.0076	-0.0015	-0.0066	-0.0014
	(0.0061)	(0.0093)	(0.0053)	(0.0023)	(0.0079)	(0.0014)
Square of Exp. in Academia	-0.0036	-0.0063	0.0003	0.0011	0.0071	0.0004
	(0.0037)	(0.0057)	(0.0016)	(0.0013)	(0.0049)	(0.0004)
Tenure-track			-0.2480			-0.0221
			(0.0054)			(0.0018)
(Research) Scientist	0.0068	-0.0010	-0.0071	0.0058	-0.0029	0.0038
	(0.0028)	(0.0039)	(0.0064)	(0.0013)	(0.0033)	(0.0020)
Engineer	0.0096	0.0102	-0.0245	0.0029	-0.0091	-0.0003
	(0.0019)	(0.0030)	(0.0060)	(0.0007)	(0.0026)	(0.0017)
Manager	-0.0047	-0.0054	-0.0300	-0.0005	0.0033	-0.0010
	(0.0020)	(0.0027)	(0.0071)	(0.0009)	(0.0023)	(0.0026)
Senior Position	-0.0036	-0.0073	-0.0486	-0.0018	0.0045	-0.0048
	(0.0017)	(0.0022)	(0.0051)	(0.0007)	(0.0019)	(0.0015)
Female	-0.0022	-0.0082	-0.0046	0.0017	0.0051	-0.0013
	(0.0028)	(0.0032)	(0.0025)	(0.0013)	(0.0028)	(0.0008)
Bachelor Matched	0.0029	0.0084	0.0103	0.0011	-0.0075	-0.0004
	(0.0030)	(0.0039)	(0.0038)	(0.0014)	(0.0033)	(0.0012)
Bachelor in the U.S.	-0.0128	-0.0143	-0.0171	-0.0048	0.0095	-0.0011
	(0.0036)	(0.0049)	(0.0044)	(0.0016)	(0.0042)	(0.0013)
Bachelor Top in the U.S.	0.0044	0.0005	0.0045	0.0018	-0.0010	0.0013
	(0.0034)	(0.0047)	(0.0033)	(0.0013)	(0.0042)	(0.0010)
N	222104	62145	121116	222153	62147	121124
Adjusted R2	.1033711	.0156813	.1068166	.0297613	.0107505	-.0328651

Notes: This table presents OLS estimates of equation 5.1. Estimates on main controls are displayed in this table, except for a missing category for gender, and for bachelor degree. See notes under Table 5.

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

	Move between Firms			Move into Top Firms		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
(Any Pub, Any Patented Pub) at t						
$d_{it}(1,0)$	0.2667	0.0392	0.1038	0.5109	0.0021	0.3440
	(0.0388)	(0.0652)	(0.0318)	(0.0816)	(0.0040)	(0.1081)
$d_{it}(1,1)$	0.1251	-0.0082	0.1292	0.2760	0.0023	0.5836
	(0.0643)	(0.0811)	(0.1077)	(0.1250)	(0.0055)	(0.2917)
(Any Pub, Any Patented Pub) between $[t-3, t-1]$						
$D_{it}(1,0)$	0.0151	-0.0047	0.1241	0.1618	0.0014	0.5996
	(0.0286)	(0.0477)	(0.0280)	(0.0561)	(0.0030)	(0.1066)
$D_{it}(1,1)$	0.1314	0.0445	0.0261	0.3530	0.0044	0.7136
	(0.0446)	(0.0658)	(0.0746)	(0.0924)	(0.0047)	(0.2343)
Controls W_{it}						
Experience = Yrs since PhD /10	0.0458	0.4059	0.2169	0.1037	-0.0241	0.7215
	(0.0510)	(0.1235)	(0.1508)	(0.1417)	(0.0079)	(0.5312)
Square of Experience	-0.2275	-0.3994	-0.5562	-0.4377	0.0218	-1.2301
	(0.0407)	(0.1009)	(0.0999)	(0.1202)	(0.0068)	(0.3520)
Cube of Experience	0.0496	0.0771	0.1366	0.0665	-0.0047	0.2678
	(0.0087)	(0.0222)	(0.0193)	(0.0276)	(0.0016)	(0.0691)
Yrs in Academia since PhD /10	0.0244	0.0755	-0.0912	-0.1239	-0.0069	-0.2179
	(0.0627)	(0.1597)	(0.1216)	(0.1636)	(0.0082)	(0.4707)
Square of Exp in Academia	-0.0662	-0.1425	-0.1426	0.0938	0.0074	-0.0308
	(0.0451)	(0.1283)	(0.0479)	(0.1110)	(0.0051)	(0.1689)
Tenure-track			-1.9690			-2.3222
			(0.0406)			(0.1593)
(Research) Scientist	0.0568	-0.0137	-0.1240	0.2646	-0.0031	0.1632
	(0.0233)	(0.0530)	(0.0519)	(0.0577)	(0.0035)	(0.1700)
Engineer	0.0861	0.1512	-0.2277	0.1824	-0.0096	-0.0510
	(0.0165)	(0.0429)	(0.0579)	(0.0440)	(0.0027)	(0.2002)
Manager	-0.0475	-0.0761	-0.2534	-0.0359	0.0035	0.0622
	(0.0176)	(0.0411)	(0.0804)	(0.0491)	(0.0025)	(0.3143)
Senior Position	-0.0324	-0.1017	-0.3886	-0.0876	0.0048	-0.3271
	(0.0143)	(0.0335)	(0.0521)	(0.0365)	(0.0020)	(0.1761)
Female	-0.0216	-0.1220	-0.0498	0.0873	0.0053	-0.1189
	(0.0224)	(0.0500)	(0.0325)	(0.0525)	(0.0030)	(0.1212)
Bachelor Matched	0.0218	0.1172	0.1266	0.0609	-0.0079	-0.0003
	(0.0230)	(0.0514)	(0.0418)	(0.0576)	(0.0035)	(0.1708)
Bachelor in the U.S.	-0.1084	-0.2089	-0.2287	-0.2902	0.0100	-0.2335
	(0.0297)	(0.0705)	(0.0488)	(0.0810)	(0.0044)	(0.1983)
Bachelor Top in the U.S.	0.0380	0.0042	0.0837	0.1542	-0.0010	0.2498
	(0.0295)	(0.0758)	(0.0432)	(0.0817)	(0.0044)	(0.1635)
N	168880	61982	79369	91788	62147	28516
pseudo R2	.1190264	.0332885	.1900324	.1684354	.0003529	.2085223

Notes: This table presents Poisson regressions of the mobility outcomes (indicators) on the same controls and fixed effects as specified in equation 5.1. The coefficients on d_{it} or D_{it} represent proportional increase in job mobility relative to the base $(0,0)$ group without a research paper.

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

	Move between Firms			Move into Top Firms		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
(Any Pub, Any Patented Pub) at t						
$d_{it}(1,0)$	0.0357	-0.0001	0.0066	0.0109	0.0031	0.0009
	(0.0062)	(0.0050)	(0.0028)	(0.0035)	(0.0041)	(0.0009)
$d_{it}(1,1)$	0.0345	0.0065	0.0090	0.0134	-0.0028	0.0017
	(0.0112)	(0.0065)	(0.0073)	(0.0064)	(0.0053)	(0.0020)
(Any Pub, Any Patented Pub) between $[t-3, t-1]$						
$D_{it}(1,0)$	0.0084	-0.0045	0.0051	0.0007	0.0041	0.0023
	(0.0044)	(0.0041)	(0.0026)	(0.0023)	(0.0035)	(0.0010)
$D_{it}(1,1)$	0.0364	0.0131	-0.0011	0.0116	-0.0022	0.0022
	(0.0083)	(0.0063)	(0.0057)	(0.0046)	(0.0053)	(0.0018)
Mean	.1177022	.067139	.0701839	.01602	.9505832	.0052813
Mean | No Pub	.1168283	.0661622	.0730854	.0153341	.950807	.0050683
Mean | Any Pub	.1543899	.0753632	.057431	.0448183	.9486985	.0062174
N	222104	62145	121116	222153	62147	121124
Adjusted R2	.1033711	.0156813	.1068166	.0297613	.0107505	-.0328651

Notes: This table presents regression estimates of equation 5.1. In contrast with Table 5, the regressions here absorb person fixed effects in addition to firm-year fixed effects. Please see notes under Table 5 for details on other controls.

Table C3: Adding Controls for Other Patent Applications

	Move between Firms			Move into Top Firms		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
(Any Pub, Any Patented Pub) at t						
$d_{it}(1,0)$	0.0353	0.0033	0.0057	0.0190	0.0033	0.0016
	(0.0059)	(0.0047)	(0.0023)	(0.0035)	(0.0039)	(0.0009)
$d_{it}(1,1)$	0.0183	0.0001	0.0045	0.0117	0.0026	0.0035
	(0.0102)	(0.0062)	(0.0064)	(0.0061)	(0.0054)	(0.0027)
(Any Pub, Any Patented Pub) between $[t-3, t-1]$						
$D_{it}(1,0)$	0.0016	-0.0001	0.0078	0.0052	0.0004	0.0040
	(0.0036)	(0.0033)	(0.0021)	(0.0018)	(0.0030)	(0.0008)
$D_{it}(1,1)$	0.0169	0.0045	0.0004	0.0122	0.0026	0.0037
	(0.0068)	(0.0052)	(0.0044)	(0.0042)	(0.0046)	(0.0020)
Other Patent Applications						
New Patent App at t	-0.0056	-0.0002	-0.0029	0.0005	0.0049	0.0040
	(0.0025)	(0.0030)	(0.0045)	(0.0011)	(0.0026)	(0.0017)
Any Patent App $[t-3, t-1]$	0.0060	-0.0031	0.0047	0.0029	0.0045	0.0014
	(0.0021)	(0.0026)	(0.0031)	(0.0010)	(0.0023)	(0.0011)
Mean	.1176511	.0672265	.0701839	.0164957	.9477359	.0060938
N	222173	62135	121116	222222	62137	121124
Adjusted R2	.1032447	.0157462	.1068225	.0330318	.0112109	-.0173464

Notes: This table presents OLS estimates of equation 5.1, but adding controls for any patent application by a worker that are not matched to her papers. Other controls and fixed effects remain the same as discussed under Table 5.

Table C4: Changes in Job Titles on Publications & Patents (Baseline Specification 5.1)

	On New Outputs d_{it}		On Lagged Outputs D_{it}	
Estimates	β_{10}	β_{11}	γ_{10}	γ_{11}
Dep. Var: Becoming a Scientist				
non-top	0.0183	0.0263	0.0122	0.0188
	(0.0053)	(0.0100)	(0.0035)	(0.0076)
Top	0.0012	0.0031	0.0020	0.0009
	(0.0038)	(0.0052)	(0.0040)	(0.0044)
Dep. Var: Becoming an Engineer				
non-top	0.0003	-0.0088	-0.0051	0.0082
	(0.0047)	(0.0081)	(0.0034)	(0.0061)
Top	-0.0088	-0.0075	-0.0090	-0.0105
	(0.0038)	(0.0048)	(0.0031)	(0.0038)
Dep. Var: Becoming a Manager				
non-top	0.0068	0.0144	-0.0090	-0.0105
	(0.0031)	(0.0063)	(0.0020)	(0.0040)
Top	0.0049	-0.0015	-0.0019	0.0065
	(0.0033)	(0.0046)	(0.0021)	(0.0038)
Dep. Var: Becoming a Scientist | Stayers				
non-top	0.0103	0.0167	0.0093	0.0183
	(0.0052)	(0.0101)	(0.0034)	(0.0078)
Top	-0.0039	0.0009	0.0035	-0.0024
	(0.0033)	(0.0045)	(0.0038)	(0.0038)
Dep. Var: Becoming an Engineer | Stayers				
non-top	0.0006	-0.0037	-0.0033	0.0035
	(0.0029)	(0.0041)	(0.0019)	(0.0038)
Top	-0.0054	-0.0044	-0.0041	-0.0056
	(0.0028)	(0.0033)	(0.0023)	(0.0029)
Dep. Var: Becoming a Manager | Stayers				
non-top	0.0070	0.0117	-0.0013	0.0033
	(0.0028)	(0.0055)	(0.0016)	(0.0034)
Top	0.0056	0.0022	-0.0038	0.0073
	(0.0031)	(0.0045)	(0.0020)	(0.0036)

Notes: This table presents estimates of 5.1 for additional promotion outcomes. (β_{10}, β_{11}) capture the responses to new innovation outputs, without or with a matched patent, respectively. $(\gamma_{10}, \gamma_{11})$ capture the responses to lagged innovation outputs, without or with a matched patent, respectively. The regression of "Becoming a Scientist" on research outputs is estimated on workers who are not a research scientist based on job titles. The second set of regressions is further restricted to stayers who remain at the same firm.

Table C5: Future Productivity

	Papers in the Next 3 years			Patented Paper in the Next 3 years		
	(1) Nontop	(2) Top	(3) Academia	(4) Nontop	(5) Top	(6) Academia
(Any Pub, Any Patented Pub) at t						
$d_{it}(1,0)$	1.7114	1.6791	1.8060	1.4333	1.4705	1.5783
	(0.0586)	(0.0510)	(0.0283)	(0.0773)	(0.0599)	(0.0663)
$d_{it}(1,1)$	1.9600	1.9740	2.1808	2.0511	2.1239	3.1825
	(0.0654)	(0.0770)	(0.0666)	(0.0769)	(0.0933)	(0.0974)
Move $j(i,t+1) \neq j(i,t)$	-0.2252	-0.5024	-0.0929	-0.5831	-1.2933	-0.1215
	(0.0583)	(0.0979)	(0.0473)	(0.0956)	(0.1597)	(0.1309)
$d_{it}(1,0) \times$ Move	-0.4396	-0.3535	-0.2212	-0.4066	-0.1116	-0.0474
	(0.1188)	(0.1513)	(0.0655)	(0.1674)	(0.2543)	(0.1885)
$d_{it}(1,1) \times$ Move	-0.5587	-0.3931	-0.4749	-0.7791	-0.1320	-0.7555
	(0.1500)	(0.1654)	(0.1776)	(0.1966)	(0.2480)	(0.2967)
N	118713	61829	104156	79970	59601	65442
pseudo R2	.4288039	.4208913	.4543796	.4573991	.4678178	.4202209

Notes: This table presents estimated Poisson regressions of future research productivity on research outputs d_{it} interacted with an indicator for any move between employers. The estimates are summarized in Table 8 as a test for model prediction 3 regarding the productivity differences between movers and stayers.

D. Estimation

D1. Estimation Procedure

Here I provide computational details on the backward induction step for solving optimal contracts given model parameters Γ (Step 1 in Section 6.1.3).

Denote by I the public information, and by $\widetilde{I}$ the private info at an incumbent employer.[72]

$t = 10$: for any potential information set $\left(I, \widetilde{I}\right)$,

- find $\left(w_{Tj}^{(1)}(\widetilde{I}_j; \Gamma), \tau_{Tj}^{(1)}(\widetilde{I}_j; \Gamma)\right)$ for incumbents (10.6), and $\left(w_{Tj}^{(0)}(I; \Gamma), \tau_{Tj}^{(0)}(I; \Gamma)\right)$ for new workers (10.11) via fixed-point iterations;
- record the continuation value $\{v_{Tj}(\cdot; \Gamma)\}$ from incumbent workers who would stay.

$t = 2, ..., 9$: given continuation values $\{v_{(t+1)j}^{(1)}(\cdot; \Gamma)\}$, find $\left(w_{tj}^{(1)}(\widetilde{I}_j; \Gamma), \tau_{tj}^{(1)}(\widetilde{I}_j; \Gamma)\right)$ for incumbents (3.12, 3.13), and $\left(w_{Tj}^{(0)}(I; \Gamma), \tau_{Tj}^{(0)}(I; \Gamma)\right)$ for new workers (3.16, 3.17) via fixed-point iterations;

$t = 1$: $\forall \pi$ (common prior), given continuation value $v_{2j}^{(1)}(\cdot; \Gamma)$, find $\left(w_{1j}(\pi; \Gamma), \tau_{1j}(\pi; \Gamma)\right)$ as in (10.18).

As shown in Section 3.2, optimal contracts depend on the labor supply $\{p_{tj}(\cdot; \gamma)\}$, and the employer-specific labor supply in turn depends on the optimal contracts. Iterations continue until we reach the fixed point

$$\bigcup_{t>1} \{p_{tj}^{(1)}(\widetilde{I}; \gamma), p_{tj}^{(0)}(I; \gamma)\} \cup \{p_{1j}(\pi; \gamma)\}$$

for any public belief π on a grid, and any origin $j \in J$ at $t > 1$.

D2. Additional Estimation Results

---

[72]Information is symmetric between employers when $y = 0$, asymmetric if $y = 1$ (3.5,3.6).

Figure D1: Profile Likelihood under Other Nuisance Parameters
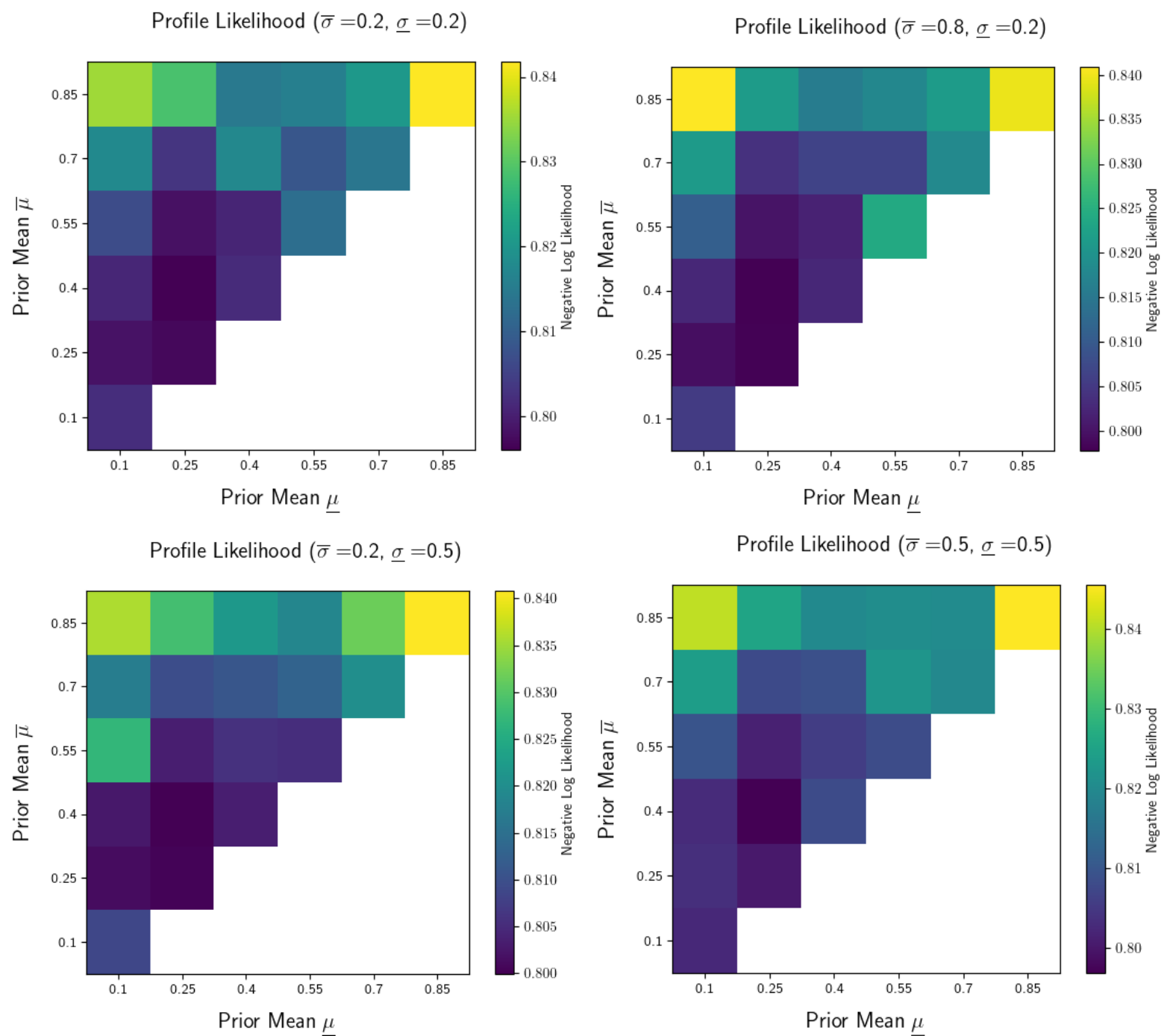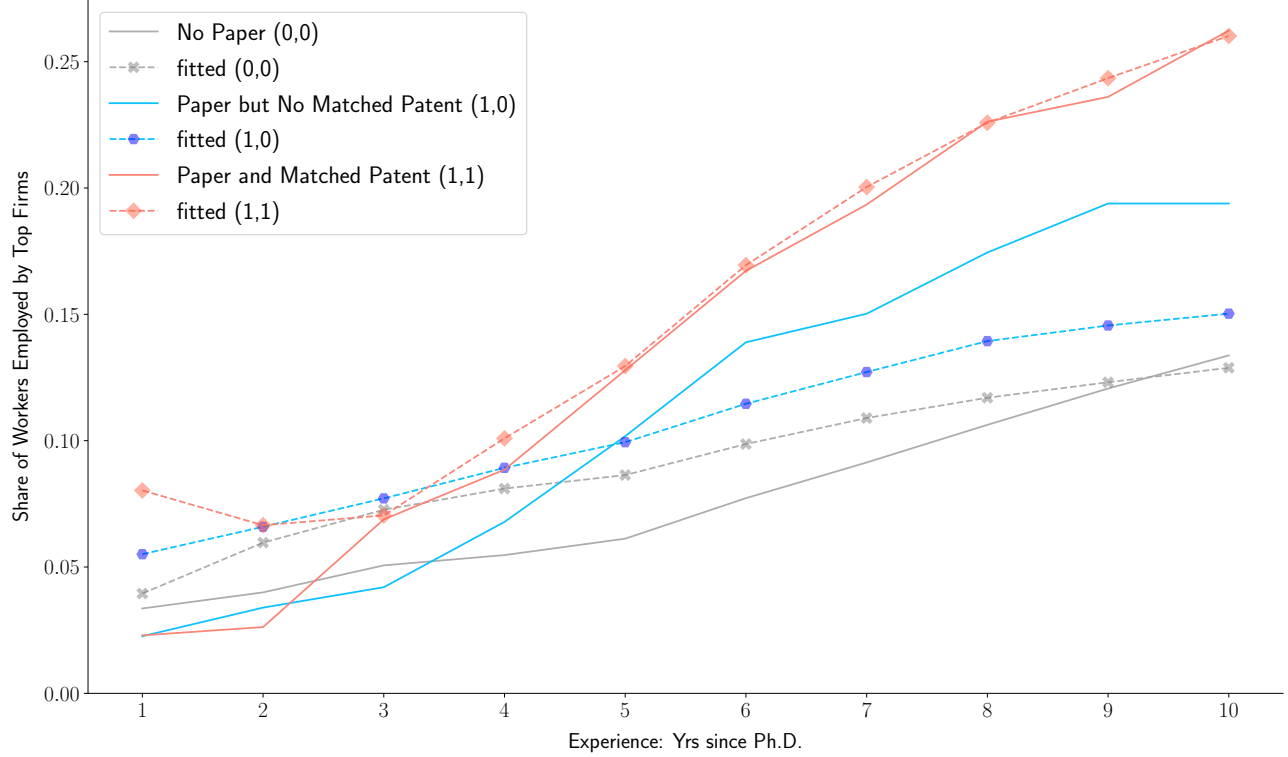
Figure D2. Upward Mobility from non-top to Top Firms

Notes: This figure shows the fitted vs. actual share of workers who are employed by one of the top firms in each year post PhD, for three groups of workers who are outside the top firms at some point in the first five years but produce different research outputs. The dashed lines are actual allocation in data, as shown in Figure 1. The solid lines are fitted shares averaged across 100 simulations of career paths, given the parameters in Table 9. See Section 6.2 for details.

Figure D3. Equilibrium Wages and Task Allocation vs. Firms' Innovation Productivity

(a) Wage Returns to Market Belief π_i versus Firm's Innovation Productivity
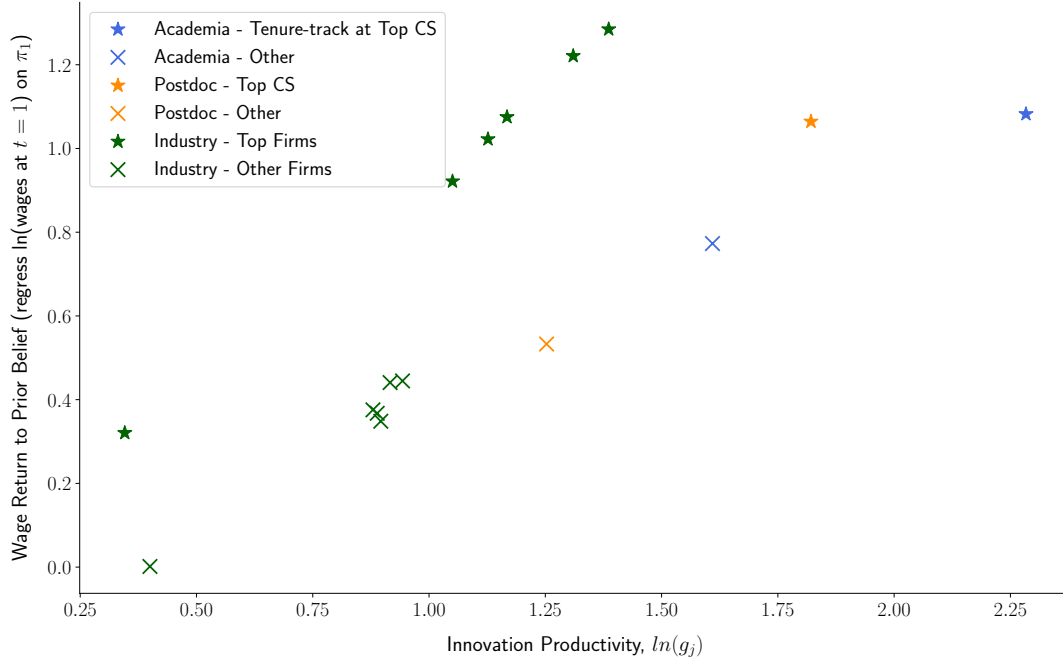
(b) Allocation to Innovation Tasks versus Firm's Innovation Productivity
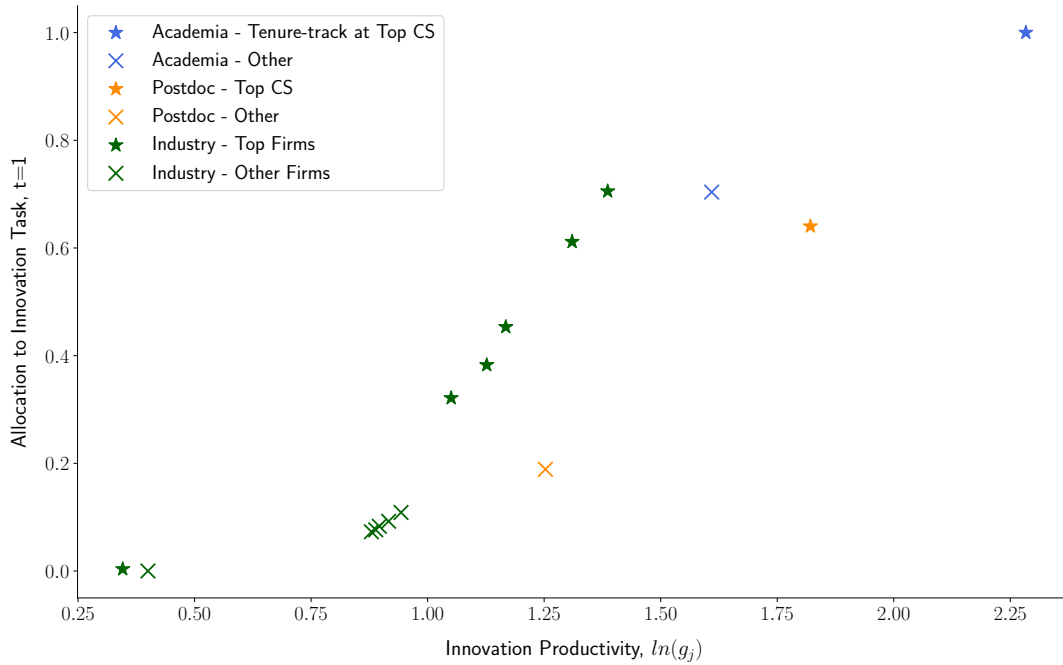
Notes: This figure shows the relationship between (a) wage returns, (b) task allocations set by firms at $t = 1$, and their (log) innovation productivity, $ln(g_j)$. Equilibrium wages and task allocations are solved via fixed-point iterations as outlined in Section 6.1.3, given the estimated parameters in Table 9. The wage returns to market belief at $t = 1$ are computed as the firm-specific regression coefficient of log wages on belief π_{i1} that a worker is high-ability.

Figure D4. Differences in Research Production

(a) Any Paper with a Matched Patent
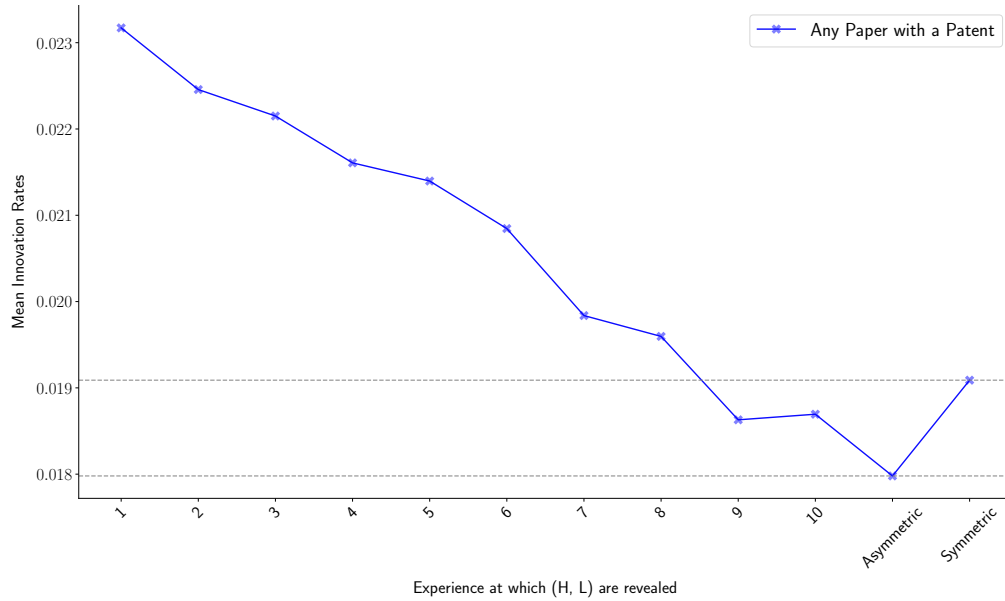
(b) Any Paper with a Matched Patent, pct change rel. to Asymmetric Benchmark
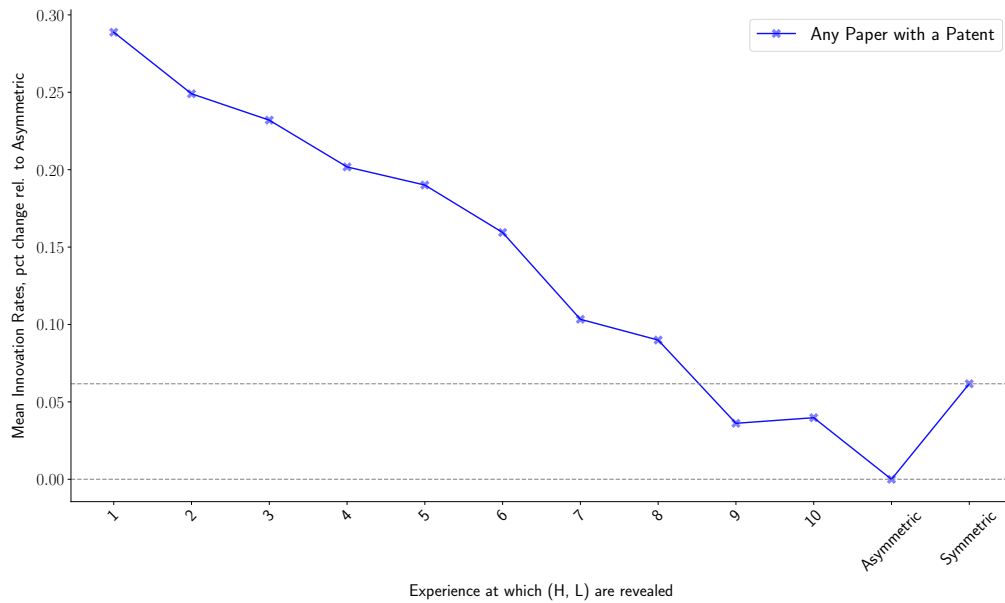
Notes: This figure shows the average rate at which workers produce a paper with a matched patent under different information disclosures. For each $t \in \{1, ..., 10\}$, (a) shows the outcome when ability α_i of a worker is fully disclosed to all employers at t. For comparison, I show the outcome under asymmetric benchmark and the counterfactual outcome under symmetric information disclosure. (b) shows the percent change relative to the asymmetric benchmark.

Table D1: Calibrated Mixing Weights

# Workers	PhD from Top 25 CS?	First Job in Research?	First Nest G_{i1}	Mean Mixing Weights $Z = E[z_{i1}]$
686	0	1	Academia - Tenure Track	0.194
1102	1	1	Academia - Tenure Track	0.327
501	0	1	Academia - Postdoc	0.126
1050	1	1	Academia - Postdoc	0.199
465	0	0	Industry - Top Tech	0.163
40	0	1	Industry - Top Tech	0.175
1063	1	0	Industry - Top Tech	0.237
114	1	1	Industry - Top Tech	0.289
2572	0	0	Industry - non-top Firms	0.022
433	0	1	Industry - non-top Firms	0.109
4062	1	0	Industry - non-top Firms	0.036
741	1	1	Industry - non-top Firms	0.155

Notes: Roughly 20% workers are labeled as $\hat{H}$ based on publication records and job titles 10 years after PhD. I use the share of $\hat{H}$ workers in each initial bin as the mean of mixing weights. The initial bins are defined based on if a worker graduates from a top 25 CS PhD programs, if her first job is in research (including all academic jobs and research scientist jobs in the industry), and the first nest G_{i1}.