The Inverted Job Ladder in Skilled Professions

Rustin Partow *

University of California at Los Angeles

JOB MARKET PAPER

This version: November 12, 2019

[Please find the latest version here]

Abstract

How do workers initially match with firms, and how do these matches improve over time? The large literature devoted to this question proposes a *job ladder* hypothesis in which poached workers move to better firms while displaced workers move to worse firms. This paper shows that the job ladder in the skilled professions is inverted, with downward-directed poaching and upward-directed displacement. I provide empirical evidence for this using a new historical dataset on lawyers. Guided by the evidence, I develop a model of dynamic labor market assignment that explains why skilled professions have an inverted job ladder. Each firm's comparative advantage is to hire a worker whose location in the talent distribution matches its own location in the job ladder. Firms' privately learn how talented their workers are and only allow below-average worker types (lemons) to be poached. Hence, poached workers move down the ladder to firms that are more specialized in lemons. Workers are revealed to have been *under-placed* when they are retained. Thus, by temporarily removing the lemons problem, random displacements help under-placed workers move up the ladder. I structurally estimate the model in order to quantify misallocation and appraise potential labor market reforms. I estimate that more than 20% of output is lost to misallocation induced by informational frictions. I find support for the use of academic competition as a means of generating stronger pre-job market signals of talent in order to reduce misallocation.

*Website: https://rustinpartow.github.io. E-mail address: rustin.partow@gmail.com.

I am deeply indebted to my advisors, Moshe Buchinsky and Maurizio Mazzocco. I also thank the following colleagues for useful conversations: Moritz Meyer-ter-Vehn, Dora Costa, El Hadi Caoui, Bernardo Silveira, Simon Board, Diego Zuñiga, Alexandre Fon, and Victoria Barone. All errors are my own.

1 Introduction

Within skilled professions, the better firms tend to work on bigger and more complex projects. DeSimone engineers skyscrapers, Cravath advises Fortune 500 mergers, and Goldman Sachs underwrites initial public offerings. The high stakes of their projects creates a need for the top firms to minimize failure by exclusively hiring top talent—a need which, due to noise in academic credentials, is met imperfectly at the initial recruitment stage. Efficiently matching skilled professionals to work that is commensurate with their talents relies on the market's ability to continually learn how talented people are and reallocate them accordingly.

The main contribution of this paper is to answer the following question: how do skilled professionals initially match with firms, and how do these matches evolve over time? The dominant framework for studying labor reallocation is known as the job ladder, whose *rungs* correspond to a ranking of firms. According to this literature, the health of the labor market is largely captured by the speed at which workers are reallocated up the ladder. The central hypothesis of the job ladder literature is that job-to-job moves tend to push workers up the ladder, while displacements into unemployment eventually lead down the ladder upon reemployment. I use a combination of theory and evidence to explain that this hypothesis is unlikely to hold in the skilled professions, which are best described by an *inverted* job ladder.

I do this in three steps. In the first step, by developing a new historical dataset on the US legal services industry, I document that job-changing lawyers move down the ladder (instead of up) to worse firms, while lawyers displaced by the dissolution of their firm move up the ladder (instead of down) to better firms. In the second step, I present a model that explains how an inverted job ladder arises from asymmetric employer information. In the third step, I structurally estimate the model in order to quantify the degree of misallocation caused by informational frictions, and to appraise potential labor market reforms.

I developed the data used in the first step by linking together annual editions of the *Martindale-Hubbell* professional directories of lawyers. The data are a comprehensive panel of all US lawyers from 1931 to 1963. Law is a particularly useful industry for exploring the inverted job ladder, and its implications, because firms can be straightforwardly ranked based on the quality of the schools that they recruit from. Law also happens to be one of the most prominent and well studied skilled professions. I use this unique dataset in order to establish several key facts on how lawyers match with and reallocate across law firms. First, I document a strong tendency of firms to specialize in distinct levels of worker talent, captured in the propensity of larger firms with better trained lawyers to recruit from better law schools. Second, after ranking firms by size and employee law school quality, I study lawyers' change in rank based on whether they appear to have been poached versus when they appear to have been displaced by shocks (defined as when their original firm exited the market). I find the inverse of the standard job ladder finding: poached lawyers move to worse firms, while displaced lawyers reemploy into better firms. Third, in order to shed some light on why poached lawyers move to worse firms, I show evidence that poached lawyers are negatively selected on unobserved talent. To do this, I leverage a unique feature of my data and setting. Lawyers with more than ten years of experience would be reviewed by Martindale Hubbell for a prestigious legal ability rating, which was eventually obtained by about a quarter of the lawyers in each geographical market. Using their future rating outcome as a latent indicator of talent, I document that retained workers are positively selected on unobserved talent.

Inspired by the above facts, I present a model of dynamic assignment that explains why skilled professions have an inverted job ladder. Firms are ranked by the difficulty of their projects. Each firm's comparative advantage in employing talent commensurate with its rank. If workers' talents were immediately known, then the model would feature immediate sorting by comparative advantage. Instead, each worker enters the labor market with an initial *resumé* of academic feats, which imperfectly captures her underlying talent and determines her initial placement. The initial or *incumbent* employer then privately learns her true talent from observing her at work. When rival firms subsequently attempt to poach the worker, the incumbent makes a counteroffer, leading to a classic lemons problem (Akerlof, 1970) where the worker is only poached if she is privately known to be of below average talent. This does not necessarily shut down all poaching. But it does mean that any firm that finds it attractive to poach from the incumbent must have a comparative advantage in its lemons, and will therefore be lower on the ladder.

Each worker's resumé evolves over time, reflecting strategic inference made by the market from a public history of poaching and retention. When a worker is retained, her resumé immediately improves, and it is revealed that she is probably under-placed at the incumbent firm. The lemons problem creates a tension in which under-placed workers cannot move up the ladder, despite the fact that doing so would raise their average productivity. Exogenous shocks that displace the worker from the firm temporarily remove this tension by eliminating the lemons problem, thus explaining why displaced workers move up the ladder.

In addition to explaining the inverted job ladder, the model provides a tractable quantitative framework for evaluating allocational efficiency and assessing potential labor market reforms. In step three, I structurally estimate the model and conduct counterfactual analysis. I compare market output to a full-information benchmark where talent is perfectly observed and fully assortative matching immediately ensues. I find that the market is 80% as efficient as the ideal benchmark, but 14% *more* efficient than a setting where all learning is shut down. Thus, the market's dynamic accumulation of information, via the evolution of workers' resumés, appears to be very important to overall efficiency.

I divide misallocation into an informational and non-informational component. The non-informational component calculates the increase in efficiency if, at each point in a representative worker's career, a social planner equipped with the same information as the market was allowed to optimally reallocate her. This essentially captures the cost of retained workers not being able to move up the ladder until the firm is dissolved. The remaining efficiency gap is informational, because it reflects the cost to the social planner of imprecise information about the worker's true talent. I find that both components of the efficiency gap are large. The informational component occurs because learning shuts down when there is still a high return to information. Once a worker's resumé reaches a certain degree of precision, production complementarities become too weak to overcome the lemons problem, and the informative (i.e., endogenous) component of turnover goes to zero. The non-informational component is subtle. Workers who are retained and revealed to be under-placed become stuck with their initial employers for too long because of the lemons problem. To compensate for this, they *over-place* when matching with the initial employer. This distortion in assignment where workers are initially over-placed and subsequently become under-placed would be eliminated in an environment with symmetric rather than private learning.

I then conduct a counterfactual policy analysis that explores the social value of a labor market reform that uses competition within the educational system to create stronger signals about talent. This analysis is intended to illustrate possible downsides of an increasing prevalence of internships as a pathway to full-time job placement in the skilled professions. Internships can undermine academic competition by allowing top students to secure a full-time job offer before completing their schooling, which disincentivizes their subsequent effort in school and thus erodes the signaling value of academic performance even for those students who did not complete internships. Thus, the rise of internships may reflect a substitution away from public signals created during school to private information collected by firms, which my model predicts will have large efficiency costs. I imagine a ban on such internships, which I model as simply delaying workers' entry into the labor market in order to make them compete against each other academically. This competition generates a binary win-loss outcome that is designed to be identical to the cutoff-style information generated by retention and poaching in the labor market. The cost of the competition is that, unlike labor market experience, it produces no output. The benefit is that at the end of the competition, the worker is unrestricted by the adverse selection problem. I find that such a policy is likely to be welfare-enhancing under reasonable assumptions about interest rates, administrative costs, and the duration of the competition.

The job ladder hypothesis has been repeatedly confirmed by aggregate labor market data, but it still has some empirical shortcomings that are directly addressed here. First, a job ladder model is typically not able to explain why a large fraction of job-to-job mobility appears to be downward directed—except by appealing to idiosyncratic shocks.[1] Second, Haltiwanger et al. (2018) recently find that the evidence for a job ladder in the US economy is primarily driven by the reallocation of workers without college degrees. They find weaker evidence of a job ladder in a sample containing only college-educated workers. My paper argues that a large segment of college-educated workers are in professions where the job ladder is inverted, which helps to settle both of these empirical shortcomings.

The inverted job ladder paints a very different picture of the overall reallocation process. Efficiency is constrained by information frictions rather than search frictions; firms are ranked by comparative rather than absolute advantage; and the outcomes of poaching and retention reflect the preferences and information of firms, not workers. The inverted job ladder should provide a better description of reallocation in labor markets where there are large returns to assortative matching between firm and worker types, and where firms acquire private information about their workers. These features seem to apply to most of the skilled professions, such as law, consulting, finance, marketing, accounting, and engineering.[2] The skilled professions (defined as occupations requiring specialized intellectual skills) are important because they hire a large and increasing share of society's brightest individuals. For example, more than a third of Harvard's undergraduate alumni accept jobs in finance or consulting alone Franck (2017).

[1]These shocks are sometimes called Godfather shocks because the poaching firm makes the worker an offer that she cannot refuse. Sorkin (2018)'s model provides an alternative explanation for apparent downward directed mobility, by essentially positing that unobserved firm-specific amenities cause some firms to be more highly ranked than they appear.

[2]Some possible counterexamples to the private information assumption would be academics and inventors, who are able to publish their work as papers or patents. However, even in these relatively unique cases, private information is likely to remain very useful in determining whether someone is likely to maintain past rates of publishing.

Having summarized the main results of my paper, I will end the introduction with a review of the related literature. The rest of the paper will be divided as follows. Section 2 describes the data. Section 3 presents the reduced form evidence on the inverted job ladder. Section 4 presents the theoretical model of the inverted job ladder. Section 5 shows how I identify and estimate the model. Section 6 performs the counterfactual analysis. The final section concludes. Proofs and technical details are in the appendix.

1.1 Related literature

Patterns in worker reallocation. This paper is related to the empirical literature on the on-the-job reallocation of workers across ranked firms, which has mostly supported a hypothesis known as the job ladder. Although the term "job ladder" was originally a generic name for a hierarchical ranking of jobs, the term now describes two stylized patterns that frequently recur in economic models of on the job search dating back to Burdett and Mortensen (1998). In a standard job ladder model, the firms that are higher on the ladder are innately more productive, are willing to pay higher wages to any given worker, and are more desirable employers. In equilibrium, a worker tends to enter the bottom of the ladder from unemployment, and gradually moves up by selectively accepting poaching offers that arrive at random. Exogenous shocks occasionally displace the worker into unemployment by destroying her current job, forcing her to start at the bottom of the ladder again when seeking reemployment. See Moscarini and Postel-Vinay (2013) for one of the latest iterations.

There is fairly abundant empirical evidence for the job ladder based on studying the transitions of poached and displaced workers using matched worker-firm data. Some recent examples of empirical evidence for the job ladder are Haltiwanger et al. (2018) using data from the Longitudinal Employer Household Dynamics (LEHD) and Moscarini and Postel-Vinay (2017) using data from the Survey of Income and Program Participation (SIPP).[3] The main factor differentiating my dataset from the datasets that have contributed to the job ladder findings is that my data come from a single insulated skilled profession.

Learning about talent. I build on the employer learning literature. The idea that asymmetric information between employers distorts mobility and impedes the efficient assignment of workers to firms comes from a long literature dating back to Waldman (1984), Greenwald

[3] The first paper ranks firms according to size, wages, or productivity, and studies net poaching outflows and inflows by rank quintile to verify that poaching is more prominent for firms at the bottom of the ladder. The second approach shows that job changers obtain relatively faster wage growth than job-stayers.

(1986), and Gibbons and Katz (1991). The main goal of this literature has been to explain empirical patterns in wages and promotions, and it has therefore emphasized heterogeneity across tasks within firms. Some examples include Bernhardt (1995), Waldman (1984), and Waldman (2016). My goal is to instead explain empirical patterns in interfirm mobility, so I focus on heterogeneity across firms. Consequently, whereas the contribution of much of the previous literature has been to explain the signaling content of job titles, the contribution of my paper will be to explain the signaling content of one's current employer. I formalize the idea that some firms are more selective than others, and thus confer different degrees of status when their names appear on resumés.[4,5]

There appears to be only one other paper that has theoretically investigated firm heterogeneity in the context of asymmetric learning: the working paper of Ferreira and Nikolowa (2019). Both of our models resolve the apparent "why do firms chase lemons" (p. 2) paradox—i.e., explain why firms poach from each other despite the winner's curse created by asymmetric learning. However, their model delivers the standard job ladder prediction of upward directed poaching. The most important difference in our models is the production function. As in the standard job ladder literature, they assume that a firm's position in the ladder indicates its absolute productivity advantage, whereas I assume that a firm's position in the ladder indicates the level of talent for which it has a *comparative* advantage.

My model features dynamic updating of beliefs about each worker's talent through an evolving resumé, which relates directly to the literature on the speed of employer learning, and presents an asymmetric information alternative to the standard symmetric learning framework of Farber and Gibbons (1996). My model is one of the first of its kind with a long time horizon, a contribution that was anticipated by the authors.

> An alternative benchmark would be "private learning," where only the worker and the current employer observe performance outcomes, but other market participants draw appropriate inferences from the observed actions of the worker and the current employer. Because the game-theoretic issues associated with such strategic information transmission can be complex, most analyses of the private-learning case have been in two period settings with special assump-

[4]Consistent with this idea, Bidwell et al. (2015) analyze survey data from investment bankers to show that higher status firms attract more talented employees without paying them more due to better signaling opportunities, which they dub the "I used to work at Goldman Sachs" effect.

[5]I choose to abstract from task heterogeneity despite the distinction between partners and associates at large firms. During the sample period of my data, only about 4% of lawyers in law firms were identified as associates. Thus, it seems reasonable to abstract from the strategic information transmission created by different job titles when using data from this period. However, it should be feasible to add this feature to my model.

> tions about functional forms and probability distributions. (Farber and Gibbons, 1996, p. 1008)

A rich literature following Farber and Gibbons (1996) has sought to test hypotheses about the nature of employer learning. The most influential prediction of these models is that over time, hard-to-observe measures of ability should become relatively more predictive of wages, and easily observed measures such as education and race relatively less predictive.[6] My model makes a similar but not identical prediction. According to my model, wages should become less correlated with ex ant characteristics like education and race and more correlated with true talent over time. However, the relationship between true talent and wages should be entirely mediated through the worker's public job history.[7]

Meanwhile, a large and more recent body of work has found evidence of private or asymmetric employer learning, where employers learn relatively more about their workers' talents than rival firms.[8] For example, Kahn (2013) estimates a model where the relative speeds of incumbent versus outside firm learning are captured by the relative variances in individual pay changes, and finds that "in one period, outside firms reduce the average expectation error over worker ability by roughly a third of the reduction made by incumbent firms." I present complementary evidence of asymmetric learning by showing that future legal ability ratings—a latent proxy for unobserved talent—are negatively predictive of current turnover, suggesting that employers selectively retain workers based on private information about their talent.

One of the most interesting contributions from this literature, starting with Altonji and Pierret (1998), has been to use the estimated speed of employer learning to indirectly assess potential justifications for schooling as a means of obtaining pre-job market signals. Most of this research makes the implicit assumption that pre-job market signaling is socially wasteful by abstracting from how information influences the quality of firm-worker matches. I find that both the private and social gains from the use of pre-job-market signaling via academic competition are large, precisely because they help workers circumvent bad initial matches, which are persistent due to the incumbent firm's private information.

[6] This implication was actually first recognized by Altonji and Pierret (2001), who confirmed it using the NLSY79 data.

[7] I don't attempt to test this prediction in the current paper because I don't have data on wages.

[8] Examples include Kahn (2013), Kahn and Lange (2014), Schonberg (2007), and Braga (2018).

2 Data and Background

My main data consist of linked entries in the annual *Martindale-Hubbell* professional directories covering US lawyers for all years between 1931 and 1963. I also match these data to deanonymized 1940 Census microdata, which I mainly use to infer permanent income from housing expenditure. *Martindale-Hubbell* (hereafter MH) is an information services company whose predecessor firms, *Martindale's* and *Hubbell's*, were founded in the mid-1800s and then merged in 1931. MH's principal products are biographical information on lawyers and legal digests.[9] Data from the MH directories have been used in several previous studies in economics and empirical legal studies.[10] MH was without a doubt the primary method for lawyers to advertise their services during the period of study.

I am aware of only one study that has attempted to transform the MH data into a comprehensive panel of individual lawyers' careers: Baker and Parkin (2006). Their paper mainly describes the process of collecting and cleaning MH's directories from 1998 to 2004, and then uses the data to describe certain new developments in the organization of law firms. Unfortunately, no additional developments appear to have come from this dataset.[11]

Each annual MH directory has a *Biographical* section, ordered by geographical markets (city/town and state), containing one or two lines of basic detail about every lawyer who responded to a questionnaire sent by MH's offices. Every person registered with the state or local bar association received a questionnaire. Professional directories like MH were the only legal method by which lawyers could advertise their services and inclusion was free, so the response rate to the questionnaires was very high.

The main purpose of the MH directory was to aid businesses searching for trustworthy lawyers in outside their usual place of business. In the early days, the legal matter at hand often involved a collection on outstanding trade credit. An excerpt from the 1902 *Martindale's* directory reads:

> The merchant would investigate with the most scrupulous care the standing of a customer before selling him a small bill of goods, but would without hesitation send a large claim for collection to a lawyer in some far-away [S]tate, of whose responsibility and trustworthiness he knew absolutely nothing; often taking a name from some one of the numerous so-called lists of "Reliable

[9] Prior to the merger, *Martindale's* had the superior biographical information, and *Hubbell's* the superior digest.

[10] Some notable examples include Garicano and Hubbard (2005), Spurr (1990), and Galanter and Palay (1993).

[11] MH seems to have become less cooperative over time in giving researchers access to their modern, computerized data.

> Lawyers," published for the purpose of advertising such lawyers, and not for the benefit of the merchant, and circulated gratis, or at a mere nominal price. Whilst this may have been excusable then, for want of other resources, it is gross carelessness now. This is the want which this work fills. It is not published in the interest of any collection agency or association, nor to advertise any special attorney or list of attorneys, but treats them all impartially, rating them as they deserve to be rated, regardless of their wishes, and is published in the interest of, and seeks its patronage from those who have business to place in their hands, thus making the very object of its existence diametrically opposite to those of any other so-called directory.

The variables that I collect from the MH directories include each lawyer's birth year, location, name, law school, the name of their law firm (if they work for one), an indicator of whether they're an associate, a legal ability rating, and an estimate of their net worth. I scraped every lawyer's entry in the MH biographical sections and then constructed a thirty-three year panel by merging individual lawyers' entries over time on the basis of their name, college, law school, and birth year. After implementing several techniques to correct for digitization errors, I was able to match about 93% of lawyers from one year to the next. To assess how much of the 7% attrition may have been caused by remaining errors, I took a random sample of 200 lawyers, aged 40-50, and manually searched for them in the directories. About 15% or 30 of the 200 cases were confirmed to be erroneous attrition caused by digitization errors that could not have been corrected by an automated procedure.[12] Thus, of the 7% attrition rate, at least 2 percentage points are caused by digitization errors.

For lawyers in law firms, a bracketed abbreviated firm name would appear beside their entry, possibly with a symbol indicating their position as an associate. The directory also contained a *Firm card* section in which firms could pay a nominal sum to advertise more details, such as who their notable clients were or the fraternal orders to which their partners belonged. I do not use this information, except to rectify a small number of firm classifications that were missing from the biographical data due to digitization errors.

The quality ratings are one of the more important and unique features of the data. MH would solicit letters from colleagues, local business leaders, and clients of each eligible lawyer and would assign to each letter a cardinal point-value. Lawyers with enough points would receive a rating ranging from *c*, *b*, or *a*. In medium to large cities, only *a*

[12]I used the panel structure of the data to try to painstakingly correct for as many of these errors as possible, and I was frequently able to correct digitization errors in year t when similar information was available in years $t-1$ and $t+1$. Unfortunately, certain individuals' names are systematically more prone to digitization error, which means that the chances of errors in two consecutive years are larger than what one might ordinarily expect.

ratings were available for only those lawyers with ten or more years of experience. The *a* ratings will be the main source of ratings data in the analysis. More details on these ratings and some of the other information is reflected in MH's confidential key in Figure 1.

1940 de-anonymized Census microdata. I match the MH data to the 100% Complete Count 1940 Census data from IPUMS in order to use expenditures on housing as a measure of permanent income in my analysis. To perform the matching, I extracted all the individuals from the Census whose listed occupation indicated a high likelihood of being a lawyer, and then used fuzzy matching on name, location, and age to match them to the MH data. If the individuals I failed to match are "unmatched at random," then dropping them from the parts of the analysis that use the Census data will not bias the results. But incorrectly matching individuals across the two datasets will bias the results, even if they are mismatched completely randomly. Because of this, I opted to leave ambiguous cases unmatched.

I successfully matched about half of the individuals in the MH data. However, this percentage is significantly higher (about 75%) for individuals spending the majority of their careers in law firms, which is the main sample of interest. One large obstacle in matching every MH lawyer to someone in the Census was that many lawyers spelled their names differently and reported slightly different birth years in the two datasets, and the resulting variations were often not sufficiently unique to make an unambiguous match. Another factor that could have prevented matching every MH lawyer to the Census is that some of the lawyers who responded to MH's questionnaire may have provided a different occupation to the Census enumerators. This could explain why lawyers working for law firms, who are likely to identify more strongly with being a lawyer, had such a higher match rate. The main application of the Census data is to identify a mapping between law school quality and permanent income, so the important question regarding selection into the sample is it obfuscates this relationship. One way to probe for this issue would be to check if lawyers from different schools were differentially selected, for which I found no evidence.

Background on sample setting. The sample period is one of relatively modest and stable growth in the legal services industry, where most lawyers worked in law firms with relatively simple transactional arrangements, leading some to dub it (Galanter and Palay, 1993) the Golden Age of Law.[13] Unlike in modern law firms, which typically feature four positions—associates, non-equity partners, equity partners, and permanent counsel—most group practice lawyers in the 1950s were identified simply as "members," or "part-

[13]This name is intended to contrast with the subsequent period of explosive growth of large law firms, beginning in the 1970s, which coincided with a greater prevalence of associates.

ners."[14]

Summary statistics on the main variables used in the analysis are included in Table 1. The sample consists of lawyer-year observations where the lawyer belongs to a law firm with three or more lawyers, is below the age of 65, and entered the market after the year 1931 (to ensure that I can see their entire career history).[15] The meanings of the *transitions* variables are described below.

Measuring mobility. Because my main interest is ranking firms and studying mobility through the ranks, I keep track of who is working with whom at each point in time, and develop a taxonomy of transitions: leaving the data (attrition), exit to sole practice, displacement, poaching, and retention. I classify lawyers into annual groups of colleagues grouping together the lawyers who are listed in the same geographical location and have the same abbreviated firm-name.[16] I refer to this grouping as a *colleague set*. Firm names are too inconsistent over time to be useful for dynamic measurements. For example, in the famous biography of one of the oldest and most prestigious law firms, known colloquially as Cravath, it is documented that the firm held six unique names in the period between 1906 and 1944 (Swaine, 1948). Therefore for the purpose of classifying interfirm mobility, I measure the similarity between colleague sets in adjacent years. Suppose that lawyer i belongs to colleague set $\mathbf{c}_{i,t}$ in year t and the $\mathbf{c}_{i,t+1}$ in year $t+1$.[17] Let C_t denote the set of *all* time t lawyers. The first measure is

$$d^1_{i,t} = \frac{||\mathbf{c}_{i,t} \cap \mathbf{c}_{i,t+1}||}{||\mathbf{c}_{i,t} \cap C_{t+1}||} = \frac{\text{Consecutive colleagues}}{\text{Time } t \text{ colleagues who stayed in the market}}.$$

The second measure is

$$d^2_{i,t} = \frac{||\mathbf{c}_{i,t} \cap \mathbf{c}_{i,t+1}||}{||C_t \cap \mathbf{c}_{i,t+1}||} = \frac{\text{Consecutive colleagues}}{\text{Time } t+1 \text{ colleagues previously in market}}.$$

In both cases, I count only the individuals who are in the sample during both time periods—otherwise, influxes of new lawyers or retirements of several older partners at once could have large effects on the results. When both of these measures are close to 1, it seems uncontroversial to assume that the firms are the same, but not when only one

[14] About four percent of lawyer-year observations in the Martindale-Hubbell data are associates.

[15] Lawyers who work alone, even if they share space and other resources with other lawyers, are sole practitioners. Using the "Class of worker" variable in the 1940 Census data, I calculated that about two-thirds of lawyers not listed in firms are truly working alone. The rest are working for the government or firms outside of law.

[16] Whereas most large modern law firms operate in multiple cities, this practice was uncommon during the sample period. In the small number of cases where firms and/or lawyers are listed in more than one location, I delete the duplicate listing in the smaller location and keep the listing in the larger location.

[17] If the lawyer has no colleagues in either year, the point is moot.

measure is close to one.[18] When the first measure is low, it indicates that the lawyer's old team does not constitute a large fraction of her new team, and it is thus likely that her team was absorbed by a larger firm. When the second measure is low, it suggests that the lawyer's old team split up.[19]

I define several nests of mutually exclusive indicators of time t worker mobility. A lawyer can continue working in group practice, exit to sole practice, or exit the dataset entirely. A lawyer exits to group practice if she is not observed in a law firm for the next two years, but remains in the sample. If there is only one intervening year of not being observed in a law firm, then she is counted as still working in group practice, and the time $t+2$ observation is used for additional classification. Given that a worker remains in group practice, she is either retained or changes jobs. A lawyer changes jobs if either distance measure is weakly below 50%.[20]

Given that a lawyer separates, she is classified as either displaced or poached.[21] A lawyer is displaced if none of her colleagues were retained. Otherwise, she is poached. Thus, displacements are intended to capture firm-wide shocks, while poaching is intended to capture mobility that is not to caused by firm-wide shocks. Although the terms poaching and displacement are perhaps more colorful than seems warranted, they are standard job ladder parlance.[22] Using these measures, I will now rank firms and study the dynamics of firm rank under the different types of mobility.

3 Empirical Evidence

This section presents empirical evidence on how lawyers are initially matched and reallocated in law. My objective is to establish three factual claims that will motivate the theory of the inverted job ladder. The first fact is that firms systematically differ in their propensity to recruit from top schools. Bigger firms with a stock of employees from better

[18]Given that law firms' main product is their talent, it would be unlikely for a law firm to re-brand while maintaining an almost identical set of employees.

[19]Despite the modest sizes of law firms during the sample period, mergers and fragmentations were somewhat common occurrences based on my readings of a large number of law firm biographies.

[20]The majority of cases are very clear-cut. A stricter or more liberal threshold would not change any of the results. However, the 0.5 threshold is preferred because it is the smallest threshold that mathematically precludes two time t colleagues who are *not* time $t+1$ colleagues from ever being counted as retained.

[21]The poaching versus displacement classification is likely to involve some error, and is only for the purpose of building qualitative evidence. In the structural estimation framework, it will be assumed that a constant fraction of separations are displacement, and this fraction will be inferred indirectly.

[22]The distinction between poaching and dissolution is only made to provide some suggestive evidence of an inverted job ladder. The structural estimation framework later on will take seriously the possibility that this distinction is made with error.

schools are likely to recruit their future employees from better schools. This fact suggests that different firms have different comparative advantages in the utilization of talent. The second fact is the inverted job ladder. Ranking firms by their size and the law school quality of their current employees, I find that poached lawyers tend to move to worse firms, while displaced lawyers tend to move to better firms. This is the inverse of the usual job ladder finding. The final fact is that poached workers are adversely selected on the basis of hidden information. The evidence comes from showing that lawyers who were retained tend to receive better future legal ability ratings than initially similar lawyers who were poached. To establish each fact, I first create a measure of law school quality.

Law school quality. By measuring law school quality, my intention is to capture an important component of a lawyer's initial perceived competence. Competence could mean analytical skills, willingness to work long hours, attention to detail, or even factors that reflect taste-based discrimination.[23] Moreover, I am only concerned with the signaling content of law school pedigree, inclusive but not exclusive to causal effects.

I construct my own continuous measure of school quality, LSQ, based on a comparison of how each school's alumni fared in three outcomes during the sample period: housing expenditure from the Census, estimated net worth scores from MH, and legal ability ratings from MH. I statistically decompose each outcome into a set of law school fixed effects after controlling for location, experience, and age. I compute each school's LSQ as the simple average of these three fixed effects, after normalizing each of them into a Z-score.[24] To corroborate the measure, I compare it to a set of ordinal law school rankings by Arewa et al. (2014).

This begs the question as to why I did not simply use the ordinal rankings directly. The problems are two-fold. First, if LSQ only had ordinal meaning, then I would be extremely limited in the types of analyses I could perform. The discussion in Section 5 will provide a theoretical foundation for using both A ratings and wealth in order to make *cardinal* comparisons across schools. Second, Arewa et al. (2014) is the most relevant ranking I have found, but even their ranking applies too much weight to recent years to be completely appropriate for my setting. It tends to overstate the quality of newer law schools, especially in the West Coast, that were still up-and-coming during my sample period.

[23]Taste-based discrimination was extremely important in 1950s corporate law firms. Corporate clients tended to be White Anglo-Saxon men listed on social registers, who preferred to work alongside lawyers from a similar background, and law firms took this into account when making hires (Swaine, 1948).

[24]I experimented with using factor analysis to choose suitable weights for the effects and found them to be very close to a simple average. The creation of the LSQ measure is more carefully described in the appendix.

Many lawyers did not attend law school early in the sample period. However, I do not have law school data for lawyers who exited the sample prior to 1939—about 17% of lawyers in the sample. For everyone else, an omitted law school should indicate that they did not attend. I treat failure to attend law school and missing law school as two separate school categories with unique LSQ measures. Most of the analysis will not use individuals with potentially missing law schools.

Fact 1: Firms Specialize in Talent. I will now show that larger firms whose existing stock of employees were trained at better schools have a propensity to recruit other employees who were also trained at better schools. To do this, I regress a lawyer's own law school quality measure on her firm's characteristics (taking care to do leave-out calculations where appropriate). I find that the propensity of firms to hire from better schools is strongly increasing in the law school quality of their existing lawyers, the share of their lawyers that have obtained A ratings, and their size. Although larger markets do in fact attract lawyers from better schools, there appears to be no residual association with market size after controlling for the aforementioned firm characteristics.

This type of sorting is difficult to rationalize without a theory where firms are comparatively advantaged in distinct levels of worker talent. Comparative advantage can arise either because of truly innate differences between firms, or because of differences in the stocks of employees those firms happen to have accumulated. In the standard job ladder literature, better firms have an absolute advantage. The economic surplus of a worker's placement is increasing with its position in the ladder, irrespective of how talented she is. This implies that the first-best assignment places every worker at the top of the ladder.

Those who are familiar with the skilled professions will recognize that the top firms are not likely to be a good fit for mediocre workers, making the absolute advantage assumption implausible. Top firms find it unattractive to hire less talented workers because their projects are more difficult and the costs associated with failure are larger. However, the absolute advantage assumption seems plausible for less skilled segments of the labor market, such as manufacturing, where firms differ in technical efficiency but not in the difficulty of their projects. he incentive for firms to specialize in distinct levels of talent may be a unique feature of the skilled professions, and helps explain why there is an inverted job ladder in law, but not in most of the labor market.

As recognized by Eeckhout and Kircher (2011) (among others), even firms that have absolute productivity advantages will only have *comparative* surplus advantages if hiring a worker prevents the hiring of someone else. In this case, the opportunity cost of hiring a worker is not just the value of her time—it also includes the foregone opportunity to hire someone who would have been a better match. Most of the standard job ladder models do not have this crowding-out effect. Again, this may be a consideration that is important in the skilled professions, where slots in firms are scarce, and other parts of the labor market.

Fact 2: Poaching leads down the ladder, displacement leads up the ladder. I will use the three measures (average LSQ, share A rated, and firm size) that I found to be highly predictive of a lawyer's own LSQ in order to rank firms and study how poached and displaced workers move through the ranks. In addition to showing the average raw changes in these variables, I will also construct a summary measure that accounts for all of them simultaneously. Specifically, I will use the regression described in Table 2 to generate a predicted value of LSQ, $\widehat{LSQ}_{i,t}$, for each lawyer i-year t observation.[25,26] I will then transform this measure into an annual percentile rank, $r_{i,t}$. Thus, changes in $r_{i,t}$ will equal the fraction of the population that the lawyer surpasses or falls behind in a given year.

In Table 3, I have computed average year-to-year changes in the different measures of firm quality, for workers who are poached, displaced, and retained. The sample excludes exit and attrition. Poached lawyers lose an average of 5 percentage points in rank, moving to firms with 7 fewer lawyers and with an average colleague LSQ that is 0.035 standard deviations lower (recall that LSQ is a standardized measure). Meanwhile, displaced lawyers experience gains that are of similar magnitude for all three measures. Retained lawyers experience relatively negligible changes—except for firm size, which grows modestly throughout the sample period. Because displacement and poaching are equally common, they tend to cancel each other out and the overall direction of mobility is essentially flat. Thus, lawyers who are poached tend to move down, whereas lawyers who seek new employment following their firm's exit tend to move up. These findings are opposite to the standard job ladder literature, where job-switching workers move to better firms, and displaced workers reemploy into worse firms.

I now analyze growth in rank conditional on current rank:

$$\Delta_{i,t} = \Delta(r_{i,t}) + e_{i,t}, \mathbb{E}[e_{i,t}|r_{i,t}] = 0.$$

I estimate $\Delta(\cdot)$ with a non-parametric kernel estimator, separately for poached, dis-

[25]To generate a more precise estimate of $\widehat{LSQ}_{i,t}$, I use a fully interacted third-order polynomial in all of the variables contained in Table 2. The estimated average changes in firm quality are not sensitive to using this rather than a first-order polynomial, but there appears to be greater dispersion in these changes when using the first-order polynomial, which seems to be expected when there is greater measurement error in the underlying firm quality that these measures are intended to proxy.

[26]$\widehat{LSQ}_{i,t}$ is not exactly the same for every lawyer in a given firm because it is a leave-out measure.

placed, and retained lawyers. For these estimates, I use the adjusted firm-rank measure.[27] The kernel estimate is plotted in Figure 2.

The estimates show that poached lawyers at high-ranking firms experience large drops in rank, while lawyers at low-ranking firms experience weak gains in rank. Job-changers departing dissolved firms obtain better average changes in rank for all levels of current rank.[28] If poaching occurred uniformly across the firm ranks, then the average poached worker would lose three percentage points in rank, while the average dissolved worker would gain about three percentage points in rank. But the inverted job ladder patterns are magnified by the fact that that poaching occurs disproportionately at the top ranks, and dissolution at the bottom ranks. In Figure 2, I plot non-parametric estimates of the share of turnover caused by each type of job change, poaching and dissolution, across the firm ranks.

It is not surprising that top-ranking firms have lower rates of dissolution. However, the fact that high-ranking firms lose relatively more of their workforce to poaching, that poached workers lose rank, and that dissolved workers gain rank, are all contrary to what we would expect from the job ladder literature.

Fact 3: Poached workers are adversely selected. Why do poached lawyers move to worse firms? To shed light on this, I will no present evidence that poached lawyers are adversely selected from those who initially joined a given firm. My data are uniquely well suited to test for evidence of adverse selection because of the availability of legal ability ratings published by MH. Because lawyers do not qualify to receive an A rating until they have 10 or more years of experience, we can think of future A rating attainment as a latent measure of current talent. I will assume that when a lawyer has between 1 and 6 years of experience, the market at this point in time does not know whether she will receive an A rating in the future. However, as the econometrician that scraped and analyzed these data more than fifty years later, I *do* know if she will receive the A rating. Thus, if poached workers have lower odds of receiving the A rating in the future, we can conclude that they are adversely selected on hidden information.

The latent-variable approach is a canonical method for testing whether firms privately learn about their employees' talents. Several papers starting with Gibbons and Katz

[27] Otherwise, it would be difficult to study the transitions of workers at the bottom ranks, because such a large fraction of their transitions are to firms where none of the colleagues reported a law school.

[28] A possible issue with interpreting these results is that the risk of firm dissolution may be related to unmeasured firm quality. If dissolving firms are unobservably worse than measured, then mean-reversion would create the appearance of downward-directed rather than upward-directed mobility following displacement. Thus, the result that job-changers from dissolved firms tend to move up the ranks seems robust to this issue.

(1991) have found evidence that workers who separated under plant closings obtained better future reemployment wages than workers who were laid off, although the adverse selection interpretation has been challenged by Krashinsky (2002), who pointed out that plant closings disproportionately affect small firms, and thus the lower future earnings of laid off workers might simply reflect disproportionate losses in size-wage premia rather than adverse selection. Some of the more recent literature tests for asymmetric learning by studying the correlation between earnings and hidden variables like, in the case of Schonberg (2007), scores on the Armed Forces Qualifying Test (AFQT).[29] Unlike AFQT scores, A ratings have not been determined at the time of the separations I consider in my test. Because A ratings are direct measures of talent, they are also not susceptible to the wage determination critique of Krashinsky (2002).

I will calculate the shares of lawyers receiving A ratings as a function of their firm's rank and their history of separation. The unit of observation is a lawyer-year, and the sample consists of lawyers who *currently* have six or fewer years of experience, and are known to remain in the sample for at least eight years.[30] In Figure 4 and Figure 5, I have plotted A ratings attainment and poaching rates by current tenure (i.e., the consecutive periods in which a worker has been retained by her original firm) for workers in the top quartile of firm rank. The results are very similar for the other quartiles, which I have also included in Appendix A.

The fact that retention is predictive of future A ratings attainment suggests that firms have private information about their employees' talent, which they use to make selective retention decisions in the face of outside options. If poached lawyers are adversely selected, then we would intuitively expect them to go to worse firms whose comparative advantage is to hire less talented lawyers. This intuition will be captured by the model.

4 A Model with an Inverted Job Ladder

I now present the model of the inverted job ladder. A worker enters the labor market with an initially sparse resumé capturing her academic credentials. Throughout the model she is assumed to not have private information about her own talent, but I show later on that this doesn't affect the qualitative results. The worker initially matches with a firm, the

[29] Two recent alternative tests of asymmetric learning are Kahn (2009) and Pinkston (2009).

[30] The first selection ensures that the A rating outcome is many years ahead, so that any shocks that are directly responsible for poaching are unlikely to directly affect the A ratings outcome. The second selection ensures that the effects of tenure are not simply picking up random exits from the data that simultaneously reduce tenure and eliminate the chance to be observed obtaining an A rating. The evidence appeared somewhat stronger without these modifications

firm privately learns her true talent, and then decides whether or not to retain her in the face of outside offers. Private learning, combined with the opportunity to match outside offers, creates a familiar adverse selection or lemons problem (Akerlof, 1970). As with the Akerlof model, incumbent firms will use cutoff rules based on private information when deciding whether or not to retain a given worker, creating a winner's curse in which only below-average workers can be poached.

The process of poaching and selective retention repeats dynamically, with occasional displacement shocks that exogenously terminate the incumbent firm and thus temporarily lift the winner's curse. The worker's resumé evolves to encode this endogenously unfolding history, which reflects some (not all) of the private information past employers held when they made their decisions.

The winner's curse does not, as might be expected, shut down all poaching. But it does mean that all rationally attempted poaching offers come from firms who are lower in the ladder than the incumbent, and which therefore have a comparative advantage in the incumbent's lemons. This is why the market exhibits downward-directed poaching. Meanwhile, when a worker is retained, her resumé improves, creating a tension in which the market knows that the worker is probably under-placed, but where the winner's curse blocks movement up the ladder. Displacement shocks resolve this tension by lifting the winner's curse, allowing under-placed workers to move up the ladder. This is the basic intuition for the inverted job ladder.

A key innovation of the model is to allow infinite horizon Markov dynamics in a private learning setting. The cutoff information structure is what keeps the model tractable. A worker's initial resumé is a simple interval describing the support of her talent, and each time a separation or retention occurs, it simply truncates the resumé from above or below. I will now describe the basic elements of the game theoretic model: the objectives of firms and workers, the timing of actions, the information structure, and the production technology.

Setup. Time $t = 1, ..., T$ is discrete and possibly infinite. There is a continuum of firms with public types $\theta \in \mathcal{R}^+$, and a single worker. The worker begins the game with a public *resumé* $[z_1, z_2]$ that signals the distribution of her talent. Her actual talent, z, is drawn uniformly from $[z_1, z_2]$ and is initially unknown to all players.[31,32] Agents seek to maximize expected earnings or profits. There is no time discounting.

[31] The assumption of one worker is equivalent to assuming that the game is separable across workers and is standard in dynamic models. The assumption of a one-dimensional, time-invariant talent or "individual competency" (Postel–Vinay and Robin, 2002) parameter is also fairly standard in the literature.

[32] Any continuous distribution can be made uniform via the inverse cumulative distribution transformation, so think of z as being the worker's percentile within some raw talent distribution.

The worker enters the game in one of two initial statuses: unattached, or attached to an incumbent firm of type θ. If the worker is initially attached, the incumbent firm privately knows her talent, z. The other firms (poachers) simultaneously make public spot-wage offers. If the worker is attached, the incumbent and the worker enter a counteroffer phase.[33]

The counteroffer phase is as follows. The incumbent privately announces a retention wage.[34] The worker decides if she wants to apply to be retained at this wage. If the worker applies, then the incumbent firm either accepts or rejects the application. If accepted, the worker receives the retention wage and works at the incumbent for an additional period. If the worker either declined to apply for retention, or was rejected, then she must accept an offer from a poaching firm. The market cannot observe whether or not the worker applied, which makes declined offers and rejections observationally equivalent.[35]

When an offer from a firm of type θ is accepted, it pays the worker the promised wage and produces output $y(\theta, z)$.[36] With probability $1 - \delta$, the worker exogenously exits the market and the game ends. Given that the game did not end, with probability $1 - \lambda^D$ the chosen firm becomes next period's incumbent. With probability λ^D, the worker is exogenously displaced from the chosen firm, and enters next period unattached. Any time a firm loses the worker, whether to poaching or displacement, it exits the game permanently.[37] However, if it is the final period $t = T$, the worker and the accepted firm produce for exactly one period, and then the game ends.

Assumption 1. $y(\theta, z)$ is twice continuously differentiable, increasing in z, strictly concave in θ. $y(\theta, z)$ is supermodular in θ and z. $y(\theta, z)$ is homogeneous of degree ϕ. Lastly,

[33]The assumption that incumbents respond sequentially to poaching offers is a common modeling choice, but for an important exception, see Li (2013).

[34]By assuming that the counteroffer is private, the incumbent will have a dominant strategy to make the minimum counteroffer required to convince the worker to apply for retention. This circumvents an analysis of whether and how an incumbent firm might use the counteroffer as a device to signal the quality of a worker who it does not intend to retain. This initially seems plausible—ex post, the incumbent firm should be indifferent to sharing information about a worker who it doesn't not intend to keep, while ex ante, commitment to such signaling would raise expected surplus. However, upon closer inspection, it seems that this would make it even more profitable for a firm to commit ex ante to "cheating" by providing exaggerated signals in states of the world where it does not intend to retain a worker. In equilibrium, such signals should become uninformative, and thus, I do not think public counteroffers would qualitatively change the results.

[35]The important thing here is that poaching firms cannot condition their offers on the worker's choice to apply for retention. Thus, poachers cannot make exploding offers—at least not the type that would contractually prevent the worker from soliciting a counteroffer within a reasonable timeframe.

[36]The price of a unit of output is exogenously set to 1, and general equilibrium changes in prices are ruled out.

[37]This assumption is similar to Assumption 9 in Bernhardt (1995), p. 319, and is made to ensure that the relevant history remains tractable.

for every z, $y(\theta, z)$ is eventually decreasing in θ.

Thus, for a given firm θ, more talent is always more productive. However, for a given level of talent z, there is a uniquely optimal firm type $\theta^\star(z)$, which, because of supermodularity, is increasing in z.[38] The homogeneity assumption is made for analytical tractability, and implies that $\theta^\star(z)$ is proportional to z. This implies that by dividing θ by the right constant, we can ensure the following normalization.

Assumption 2. $\theta^\star(z) = z$.

Definition 1. Suppose that the worker's resumé is $[z_{1,t}, z_{2,t}]$.

1. $y(z, z)$ is the *full information output*.
2. $y^{\text{FIM}}(z_{1,t}, z_{2,t}) = \frac{1}{z_{2,t} - z_{1,t}} \int_{z_{1,t}}^{z_{2,t}} y(z, z)dz$ is the worker's *expected full information output*.
3. $\bar{y}(\theta, z_{1,t}, z_{2,t}) = \frac{1}{z_{2,t} - z_{1,t}} \int_{z_{1,t}}^{z_{2,t}} y(\theta, z)dz$ is the *ex ante average output*.
4. $\bar{y}^{\max}(z_{1,t}, z_{2,t}) = \max_\theta \bar{y}(\theta, z_{1,t}, z_{2,t})$ is the *optimal ex ante average output*
   (a) $\theta^{\max}(z_{1,t}, z_{2,t})$, the maximizer associated with $\bar{y}^{\max}(z_{1,t}, z_{2,t})$, is the *optimal ex ante placement*.
   (b) If θ is larger (smaller) than $\theta^{\max}(z_{1,t}, z_{2,t})$, the worker is *over-(under-)placed*.

In general, the full information output will serve as an unattainable, first-best benchmark. If a social planner had knowledge of the resumé, but not of z, then at best she could obtain the optimal ex ante average output by allocating the worker to her optimal ex ante placement. We shall see that the worker's resumé evolves over time, and equilibrium placement will not coincide with these statically optimal placements—an important difference with exogenous public learning models. Private information prevents a currently attached worker from leaving her firm unless she turns out to be below average talent. This causes a worker who is repeatedly retained to become under-placed. To partially offset the costs of future under-placement, the worker is incentivized to initially over-place when matching with a new firm.

To summarize the timing, each period has three stages. In stage one, poachers make wage offers. Stage two features the counteroffer and application/rejection stage, if the worker is attached. If by stage three the worker is unattached, then she selects her preferred poaching offer. In order to build intuition, I will start with a one-period model and

[38]Of course, supermodularity is not sufficient for this result. It is also necessary that $y(\theta, z)$ eventually be decreasing in θ. Otherwise, $\theta^\star(z)$ would equal infinity for all values of z.

then append more periods. I will conclude with results for the infinite horizon Markov equilibrium, which is the main object of analysis.

Definition of an equilibrium. An equilibrium is a collection of (1) beliefs about the worker's talent as a function of the history of the game, (2) wage offer rules, and (3) job acceptance rules such that the beliefs are consistent with Bayes' rule, whenever it applies, and the wage offer and job acceptance rules are sequentially rational. I will now introduce a refinement on off-equilibrium-path beliefs, which is in the spirit of the Divinity Criterion of Banks and Sobel (1987).

Assumption 3 (Divinity)**.** Suppose that after some history of play, the market assigns minimum value z_1 and maximum value z_2 to the worker's set of possible talents. If, along the equilibrium path, the incumbent firm is expected to retain the worker with probability 1, but the worker is not retained, beliefs update to assigning full probability to $z = z_1$. If instead the incumbent firm was expected to retain the worker with probability zero, but she is subsequently observed to be retained, beliefs update by assigning full probability to $z = z_2$.

This assumption is intuitive in the sense that, among the set of possible levels of talent consistent with the previous beliefs, beliefs update to place full probability on the level of talent that would make the observed off-path play *most* profitable. This assumption is also necessary to ensure that beliefs are continuous functions of the equilibrium cutoff rule. Cutoff rules that approach zero or one-hundred percent poaching probabilities will generate beliefs that converge in probability to z_1 (or z_2), and so another way of stating Assumption 3 is that beliefs equal their limits.

An overview of the model. The two main implications of the model are that (1) poached workers move to firms that are lower ranked than their incumbent, and (2) displaced workers move to firms that are better ranked than their incumbent. The worker's talent will be uniformly distributed within her resumé $[z_{1,0}, z_{2,0}]$ when she initially enters the market and joins her first firm. Poachings will truncate the resumé down, retentions will truncate the resumé up, and displacements will have no effect on the resumé. Once employed at an incumbent firm, the worker is partially insulated from outside wage competition due to the winner's curse afflicting potential poaching firms. She is always made a counteroffer that makes her no better off than being poached (and perceived to be below the firm's cutoff). Thus, displacements are *always* payoff enhancing from the perspective of the worker, because they free her from the winner's curse and allow her to receive more competitive poaching offers.[39]

The one-period model. Let us quickly explore what happens in a one-period version of the model, when $t = T$. Using backward induction, we know that in stage three the worker will choose the highest wage offer. In stage two, having observed the highest poaching offer, the incumbent firm's dominant strategy is to announce a potential counteroffer that matches it. The worker's dominant strategy is to apply for the counteroffer.[40] The incumbent then decides whether or not it is worthwhile to retain the worker at the previously announced counteroffer.

Let θ_T the incumbent firm, z the worker's true, and w_T^R the retention wage. The incumbent will retain the worker if and only if $z_T > \zeta_T$ where ζ_T is the unique level of worker talent such that $y(\theta, \zeta_T) = w_T^R$. Thus, retention follows a cutoff rule. Now let us consider the poaching offers made in stage 1. Each poacher knows that if its wage offer is pivotal, it will determine the counteroffer w_T^R, and therefore the cutoff rule, to be used by the incumbent in stage 2.

Lemma 1 (The time-T cutoff)**.** Recall that $\bar{y}^{\max}(z_{1,t}, z_{2,t})$ is the maximized expected output of a worker with resumé $[z_{1,t}, z_{2,t}]$, and ζ_T is the equilibrium cutoff rule. The equilibrium cutoff rule is the largest ζ_T satisfiying

$$y(\theta, \zeta_T) \leq \bar{y}^{\max}(z_{1,T}, \zeta_T) \tag{1}$$

The equilibrium cutoff, $\zeta_T(\theta, z_1, z_2)$ is the maximum value within $[z_{1,t}, z_{2,t}]$ satisfying Equation 1.

See proof on page 57.

Equation 1 explains that along the equilibrium path the marginally retained worker's output must be equal to the average poached worker's output. However, this is merely a necessary condition, not a sufficient one. It is tempting to imagine a locus of wage-cutoff rule equilibria satisfying this necessary condition, some with endogenously low poaching, and others with high poaching, in the spirit of Acemoglu and Pischke (1998). The difference here is that poaching firms have a first-mover advantage, and will always force coordination on the highest level of poaching satisfying Equation 1.

[39] Firms would profit by being able to commit against their inefficient rent seeking behavior, for example by promising to reveal all of their private information. The model assumes that firms cannot commit.

[40] As is often the case, in equilibrium the worker must behave as if she prefers the incumbent in cases of ties, because were this not the case, then the incumbent could always pay an infinitesimally small premium to break the tie.

Lemma 2. Suppose that $z_{1,T} \leq \zeta_T(\theta_T, z_{1,T}, z_{2,T}) < \theta_T < z_{2,T}$. Then the equilibrium poaching firm is of lower type than the incumbent firm.

See proof on page 58.

Poached workers are lemons, and tend to join firms that are more specialized in lemons. These will tend to be firms that are lower on the ladder than the incumbent was. The possible exception is that if the marginally retained worker were much higher than θ (the worker type for whom the incumbent is uniquely optimal), then the lemons of the incumbent might actually be suited for a better firm. This condition will be justified by adding periods and modeling the choice of the incumbent firm.

The model with two or more periods. Now let us append more periods to the game. The worker starts the game at period $T-1$ with resumé $[z_{1,T-1}, z_{2,T-1}]$, either unattached, or attached to a firm of type θ_{T-1}. In stage 1, firms make poaching offers. In stage 2, if the worker is attached, then the incumbent posts a retention wage, the worker chooses whether to apply, and if she applies, the incumbent reveals a rejection or retention outcome. Retention causes the market to infer that the worker was above the equilibrium cutoff. Both rejection and declining to apply lead the market to infer that she was below the cutoff. In stage 3, unless the worker was retained, she chooses her preferred poaching offer.

The structure of the subgame beginning in period T will be identical to the one-period model: the worker is either unattached, or attached to an incumbent firm. Her talent is uniformly distributed along an interval (the resumé). By backward induction, any T period model can be recast into the exact same structure: poachers make outside offers, the incumbent makes a counteroffer, the worker matches with a firm of type θ, and then the players enter a continuation game that was previously solved. For the two-period model, I have depicted a graphical rendition of the game in Figure 6.

I will now define the core functions that allow me to solve for the unique equilibrium. For the sake of clarity, I will allow the game to start at different possible times, $t = -\infty, ..., T$. Thus, $t = T$ corresponds to the one-period model, $t = T-1$ the two period model, and so forth. Conditional on the worker's attachment status, resumé, and identity of incumbent firm, it is irrelevant whether a game literally started in period K, or is merely a continuation of a game that started earlier.

Definition 2. The *indirect poaching utility function*, $V_t^P(\theta, z_1, z_2)$, is the expected net present earnings of the worker who begins the game in period t unattached and with resumé $[z_1, z_2]$.

Definition 3. The *value function* is $V_T(z_1, z_2) = \max_\theta V_T^P(\theta, z_1, z_2)$.

Definition 4. The *cutoff rule function*, $\zeta(\theta, z_1, z_2)$, is the equilibrium cutoff rule used when the worker begins the game in period t with resumé $[z_1, z_2]$ attached to incumbent firm θ.

If the worker chose not to apply for retention, then as depicted in Figure 6, the market would incorrectly believe that she had been rejected, and her beliefs would diverge from her resumé. The next lemma states that a worker'(potentially divergent) beliefs are irrelevant to all future payoffs, conditional on her resumé, attachment status, and incumbent firm.[41]

Lemma 3. Conditional on the resumé, attachment status, and incumbent firm, the worker's own beliefs do not affect her payoffs, and it is a dominant strategy to apply for retention.

See proof on page 58.

The firm will pick a counteroffer makes the worker indifferent to applying and not applying. By Lemma 3, this is equivalent to making the worker indifferent between retention and rejection. It will be useful to imagine what this wage would be across different hypothetical cutoffs (which may or may not be equilibrium choices).

Definition 5. The *hypothetical retention wage function*, $w_t^R(\theta, z_1, z_2, \zeta)$, is the retention wage that makes the worker entering the game at time t with resumé $[z_1, z_2]$ attached to incumbent θ indifferent to being rejected and obtaining $V_t(z_1, \zeta)$ in the current period, and being accepted.

Lemma 4.

$$w_t^R(\theta, z_1, z_2, \zeta) = V(z_1, \zeta) - \delta \left((1 - \lambda^D) V(\zeta, \zeta(\theta, \zeta, z_2)) + \lambda^D V(\zeta, z_2) \right).$$

See proof on page 59.

The equilibrium cutoff must equate the output of the marginally retained worker to the hypothetical retention wage. It is the largest possible cutoff satisfying this condition.

Proposition 1. The *equilibrium cutoff rule* is

$$\zeta_t(\theta, z_1, z_2) = \sup_{\zeta \in [z_1, z_2]} \left\{ y(\theta, \zeta) - w_t^R(\theta, z_1, z_2, \zeta) > 0, \forall z > \zeta \right\}.$$

See proof on page 59.

Definition 6. The *retention wage function*, $w_t^R(\theta, z_1, z_2)$, is equal to $w_t^R(\theta, z_1, z_2, \zeta_t(\theta, z_1, z_2))$.

[41] Lemma 3 indeed implies that the main analysis would be unchanged by assuming that the worker also has private information about her true talent.

The output of the marginally retained worker must be weakly higher than the retention wage, and strictly so for any level of talent above the cutoff. After hiring the worker and producing output, the successful poaching firm will, except in the event of displacement, become next period's incumbent and learn the worker's talent.

Definition 7. The *ex ante incumbent profit function*, $\Pi_t^I(\theta, z_1, z_2)$, is the expected net present profits of an incumbent firm of type θ, given that the worker's true type z is uniformly distributed between z_1 and z_2.

The profits of becoming the incumbent are weakly positive because of the outside option of rejecting the worker, exiting the game, and obtaining a zero payoff. If the incumbent privately discovers that the worker is highly talented, then it may earn strictly positive profits, ex post. Thus, the ex ante profit function captures the expectation that the firm makes when it only knows the worker's resumé and has not yet acquired its private information. This expectation must account for the fact that, if a worker's resumé is $[z_{1,t}, z_{2,t}]$ and the incumbent's cutoff rule is ζ_t, then the worker's resumé conditional on being retained will be $[\zeta_t, z_{2,t}]$, and the probability of retention will be $\frac{z_{2,t}-\zeta_t}{z_{2,t}-z_{1,t}}$.

To induce the worker to apply for retention, the incumbent will announce a retention wage, w_t^R, that makes the worker indifferent between retention and rejection. If the worker is retained, then with probability λ^D she will be displaced, and with probability $1-\lambda^D$ she will remain at the incumbent firm where she will again be made indifferent to retention and rejection. The next lemma uses this insight to solve for the retention wage.

Proposition 2. The retention wage function is given by

$$\begin{aligned} w_t^R(\theta, z_1, z_2) = & V(z_1, \zeta_t(\theta, z_1, z_2)) \\ & - \delta\left(\lambda^D V(\zeta_t(\theta, z_1, z_2), z_2) + (1-\lambda^D) V(\zeta_t(\theta, z_1, z_2), \zeta_{t+1}(\theta, \zeta_t(\theta, z_1, z_2)))\right). \end{aligned}$$

The profit conditional on retaining the worker is therefore expected output in the current period, minus the retention wage, plus the future incumbent profits (appropriately discounted by the probabilities of exit and displacement). The ex-ante incumbent profit function is characterized recursively below.

Proposition 3. For all $t > T$, $\Pi_t^I(\theta, z_1, z_2) = 0$. For all $t \leq T$,

$$\begin{aligned} \Pi_t^I(\theta, z_1, z_2) = & \frac{z_2 - \zeta_t}{z_2 - z_1}\left(\bar{y}(\theta, \zeta_t, z_2) + \delta(1-\lambda^D)\Pi_{t+1}^I(\theta, \zeta, z_2) - w_t^R\right), \\ & \text{subject to } \zeta_t = \zeta_t(\theta, z_1, z_2), \\ & \quad \text{and } w_t^R = w_t^R(\theta, z_1, z_2). \end{aligned}$$

To complete the characterization of the model, I solve the indirect poaching utility. The indirect utility has three components: (1) the current wage, equaling expected output plus future ex-post incumbent profits, (2) the continuation value if the worker is not displaced, and (3) the continuation value if the worker is displaced.

Proposition 4.

$$
\begin{aligned}
V_t^P(\theta, z_1, z_2) = & \underbrace{\bar{y}(\theta, z_1, z_2) + \delta(1-\lambda^D)\Pi^I_{t+1}(\theta, z_1, z_2)}_{\text{wage}} \\
& + \underbrace{\delta(1-\lambda^D)V_{t+1}(z_1, \zeta_{t+1}(\theta, z_1, z_2))}_{\text{continuation value of adversely selected worker}} \\
& + \underbrace{\delta\lambda^D V_{t+1}(z_1, z_2).}_{\text{continuation value of displaced worker}}
\end{aligned}
$$

Analysis of the infinite horizon Markov equilibrium. As T goes to infinity, all of the equilibrium objects defined so far converge to stationary functions, and we can drop t subscripts. In Figure 7, I have illustrated the infinite horizon Markov equilibrium. For simplicity, I have ignored the possibility of displacement. When displacement occurs, the worker's continuation value is $V(z_1, z_2)$ instead of $V(z_1, \zeta)$, and her poaching firm is $\theta(z_1, z_2)$ instead of $\theta(z_1, \zeta)$.

The inverted job ladder. I will now show prove that the model predicts an inverted job ladder, with downward-directed poaching and upward-directed displacement. Assume henceforth that the worker initially enters the labor market unattached. First, I will show that, under certain conditions, the cumulative increase in $z_{1,t}$ during any particular employment spell—and thus, the cumulative poaching risk—is bounded.

Proposition 5. For every possible (z_1, z_2, θ), there is some $\bar{\delta}$ such that, if and only if $\delta > \bar{\delta}$, the incumbent's cutoff rule is z_1. $\bar{\delta}$ is decreasing in $\frac{z_1}{z_2}$.

See proof on page 59.

This proposition helps guarantee that the marginally retained worker is always below the comparative advantage of the incumbent firm. However, this would be trivially satisfied if there were no poaching at all, so the next corollary establishes no poaching is not the only case where Proposition 5 applies.

Corollary 1. There exists a range of values for δ for which there is strictly positive poach when $\frac{z_1}{z_2}$ is sufficiently small, but where $\zeta(\theta, z_1, z_2) < \theta$ at all points along the equilibrium path.

Thus, as long as the discount factor is sufficiently high, we can ensure an arbitrarily small bound on the degree to which the worker's resumé can improve throughout the employment spell. The basic intuition is that, as workers become sufficiently patient, the *stigma* cost of being revealed below the incumbent firm's cutoff becomes arbitrarily high, relative to any contemporaneous difference in wage offers. The desire to protect her resumé causes the worker to accept relatively low wages in order to remain at the incumbent firm. It may seem pointless to protect one's resumé if doing so requires remaining at a firm that is paying very low wages. But recall that there is always some probability of becoming exogenously displaced in the future, at which point wage offers will become competitive.

This means that if δ is sufficiently large, any particular employment spell beginning at (z_1, z_2, θ) has a cap on the degree to which z_1 can increment up through market updating.

Assumption 4. Assume that δ is sufficiently low to guarantee some degree of poaching, but sufficiently high to ensure that cutoff rules are below the incumbent's type at all points along the equilibrium path.

The next two theorems are the main results.

Theorem 1. Assume that $\theta(z_1, z_2)$ is increasing in z_1 and z_2. Along every equilibrium path with poaching, the poaching firm is of lower type than the incumbent firm. I.e., $\theta(z_1, \zeta(\theta, z_1, z_2)) < \theta$ for all (θ, z_1, z_2) along the equilibrium path.

Proof. By the previous assumption, $\zeta(\theta, z_1, z_2) < \theta$. In this case, the optimal firm $\theta(z_1, \zeta)$ is below $\theta(\zeta, \zeta)$, and since $\zeta < \theta$, this is below $\theta(\theta, \theta)$, which is, trivially, equal to θ. □

Theorem 2. Assume that $\theta(z_1, z_2)$ is increasing in z_1 and z_2. Along every equilibrium path with displacement, the worker reemploys with a new firm that is of higher type than her previous firm. If, in the previous employment spell, the worker was retained during one or more periods with non-zero poaching probabilities, then the new firm is of strictly higher type.

Proof. Throughout an employment spell, the lower bound of the worker's resumé weakly improves. If the worker is retained after a period with a non-zero poaching probability, it strictly improves. Therefore, upon being displaced, the worker's resumé is better (strictly) than it was when she chose to match with the firm. Because the endogenous choice of firm, $\theta(z_1, z_2)$, is increasing in the resumé, the worker will choose to match with a (strictly) higher ranked firm. □

5 Model Estimation

I will solve and simulate the infinite horizon Markov model in order to quantitatively assess the efficiency of talent discovery and reallocation in the market for lawyers, and to understand the value of potential labor market reforms. To do so, I must pick realistic model parameters. Thus, I prove that these parameters are identified from the observable data, and construct closely related estimators.

5.1 Identification

Overview of identification. For the purposes of identification, the observed data will be a lawyer-year panel containing each lawyer's years of experience, law school, home expenditure, A rating attainment, indicators for whether she exited to sole practice or exited the data, and an indicator for whether or not she separated from her firm. I do not assume that poaching and displacement are accurately distinguished in the data. I need to identify the exogenous exit and displacement rates, the production function, and the initial distribution of resumés.

I will calibrate the exit rate δ to the rate at which lawyers are estimated to exit group practice law, within a sample of lawyers below age 55.[42] I will identify λ^D by the degree to which retention positively signals future A ratings attainment. To identify the two other model features—the production function and the initial distribution of worker resumés—I make additional restrictions. I make a functional form restriction on the production function which parametrizes it into (ϕ, α), where ϕ is the return to scale in production and α captures the degree of complementarity—or, equivalently, the cost of mismatch—between firm and worker types. I also restrict heterogeneity in initial resumés to preserve a constant ratio of $\frac{z_{2,0}}{z_{1,0}} = \gamma$, where the parameter γ captures how how much information is contained in workers' initial credentials. The parameters ϕ and γ are jointly identified from a closed form relationship between the share of each law school's alumni that are A-rated and their average permanent income. I then identify α from turnover rates. The estimation methodology closely follows the identification results.

Main Assumptions. The main assumptions for identification are as follows.

[42]In general, δ is a difficult parameter to choose because, if we also wanted to capture time-preferences (which are ignored in the theory section), we would alter δ. Initially one might think that time preferences should result in an even lower value for δ. However, if we wanted to model secular trend increases in the profitability of legal services over time, that might instead suggest using a higher value for δ.

Assumption 5. Let $(z_{1,0}, z_{2,0})$ denote the initial resumé. Then $z_{2,0} = \gamma z_{1,0}$, for some fixed constant γ. There exists an observed variable, $\mathbf{x}$, such that

$$\ln z_{2,0} = g(\mathbf{x}) + \epsilon, \epsilon \text{ is independent of } \mathbf{x}.$$

Assumption 6. A lawyer eventually achieves an A rating if and only if her talent is above a fixed threshold z_A. The A rating outcome is not informative to the market.

Assumption 7. Let $\alpha \in (0, \phi)$, $\phi > 0$.

$$y(\theta, z) = \phi \theta^{\alpha} z^{\phi - \alpha} - \alpha \theta^{\phi}.$$

The first part of Assumption 5 requires that all initial heterogeneity in workers' resumés takes the form of proportional shifts, which hold $\frac{z_2}{z_1}$ constant. The second part requires an instrument for this proportional heterogeneity. This assumption, combined with Assumption 6, provides me with the structure necessary to establish a mapping between A ratings attainment and permanent income that identifies ϕ and γ. In order to coherently use A ratings without having incorporated them as additional sources of public information in the model, I assume that they are not informative to market participants. Think of this assumption as a mere approximation capturing the fact that A ratings are no revealed until year 10.

Assumption 7 is a functional form restriction. The first term in the production function can be interpreted as revenue, and the second term as an operating cost (exclusive of wage payments to the worker). Higher ranked firms and more talented workers are more productive, and more so when matched with each other. However, higher ranked firms are costlier to run, which puts them at a comparative disadvantage in employing less talented workers. Notice that the operating cost need not be an accounting cost. It could be an opportunity cost reflecting, for example, that firms have scarce slots which are more valuable at higher type firms.[43]

Assumption 7 now satisfies all of the properties of the production function described in Section 4. The choice of coefficients on the revenue and opportunity cost terms, ϕ and α, is arbitrary. Because these coefficients are generically not separately identified from linear transformations of θ and z, I have chosen the parametrization respecting the normalization in Assumption 2 that $\theta^{\max}(z) = z$. With this production function in hand, it

[43]One potential extension of my model would be to derive these opportunity costs in general equilibrium. Allowing cyclical shocks to endogenously change firms' opportunity costs would probably be the most fruitful avenue for connecting the model to the business cycle, which has been a major emphasis of the standard job ladder literature.

is possible to analytically derive the objects that I defined in Definition 1 capturing output optimization under various degrees of certainty about the worker's talent.[44]

This choice of production function features a particularly convenient parameterization of the costs of over- or under-placement.

Lemma 5.

$$\frac{\partial^2 \ln y(\theta, z)}{\partial \ln \theta^2} \mid_{\theta=z} = -\alpha\phi.$$

Thus, $\alpha\phi$ governs the percentage loss in efficiency as θ departs from z. Given an initial estimate of ϕ, attributing a larger part of ϕ to α will intuitively raise the costs of a given degree of over- or under-placement. For this reason, we should also expect that raising α to increase turnover, vis-a-vis the poaching of lemons by lower ranked firms. The effects on efficiency (expected output as a fraction of full information output) are not as intuitive. On the one hand, a higher value of α will certainly imply lower efficiency at the beginning of a worker's career. However, by increasing turnover and speeding up the pace of market inference, a higher value of α could also result in greater efficiency towards the end of a worker's career. Markets with a low α will be efficient and markets with a high α will be inefficient, but it is not completely clear whether the efficiency consequences of private learning depend monotonically on α.

Proposition 6 (Identification of λ^D)**.** Consider a lawyer with a given time t history. Let λ^D denote the rate of exogenous displacement. Assume that the probability of achieving an A rating is strictly between 0 and 1 in the event that the lawyer is retained. Then λ^D is identified by

$$\lambda^D = Pr(\text{Separate}|\text{Achieve A rating}).$$

See proof on page 60.

Intuitively, some degree of turnover is informative (poaching), and some degree is not (displacement). In a world with no exogenous displacement, separation should imply a very negative signal, and a separating worker's talent should be less than that of a retained worker with the same prior history with probability one. Thus, in such a world, A rated lawyers should *never* have experienced a separation. On the other hand, if the future A rating is completely uninformative about the likelihood of separation, and all separation must be exogenous.

Proposition 7 (Identification of ϕ and γ)**.** Let $\mathbf{x}$ denote the instrument (such as LSQ) described in Assumption 5. Suppose there exists an interval of values for $\mathbf{x}$ where the

[44] For interested readers, these are all derived in Appendix B.

probability of A ratings attainment conditional on $\mathbf{x}$ is strictly between 0 and 1. Let $\tilde{v}(\mathbf{x}) = \mathbb{E}[\ln V|\mathbf{x}]$, and let $p_A(\mathbf{x})$ denote the probability of becoming A rated conditional on $\mathbf{x}$. Let $\tilde{\gamma} = \frac{\gamma-1}{\gamma}$. Because $\tilde{v}(x)$ and $p_A(\mathbf{x})$ are both increasing, there exists a one-to-one relationship $\tilde{v}(p_A)$. Moreover, ϕ and γ are identified by

$$\phi = \left(\frac{\partial \tilde{v}}{\partial p_A}\right)^2 / \frac{\partial^2 \tilde{v}}{\partial p_A^2},$$

and

$$\tilde{\gamma} = \frac{1}{\frac{\partial \tilde{v}}{\partial p_A} / \frac{\partial^2 \tilde{v}}{\partial p_A^2} + p_A}$$

See proof on page 61.

Thus, ϕ and γ are identified from two distinct measures of curvature in the relationship between $\tilde{v}$ and p. This identification result leverages an important and unique aspect of my data. Without A ratings, there would be no way to anchor differences in ex ante characteristics, such as LSQ, into cardinal differences in talent. It would therefore be impossible to say whether a given return to LSQ was driven by large *differences* in talent between schools, or a large *return* to talent. The difference is crucial for correctly quantifying market efficiency.

Finally, I identify α from turnover.

Proposition 8. The first-year turnover rate is monotonically increasing in α, holding ϕ fixed.

See proof on page 62.

5.2 Estimation

To estimate the parameters, I restrict the sample to lawyers who are observed entering the dataset between the ages of 22 and 30 and immediately started working for a firm, and who do not switch in and out of sole practice. This restriction ensures that the job histories of the lawyers in the sample are sufficiently similar to make Assumption 5 plausible. The annual sample attrition rate is estimated to be 2.12% for lawyers below the age of 55, and 2.77% for lawyers above the age of 55, leading me to assume that the difference—0.65%—is the true rate at which lawyers exit the data, while the rest is erroneous attrition. Lawyers in the sample exit to sole practice at a rate of 6.22%. Thus, δ is calibrated to equal $1 - 6.87\% = .931$.

The rate at which lawyers exit into sole practice is estimated to be

Estimation of λ^D. To estimate λ^D, I use the previous result stating that $\lambda^D = Pr(\text{Separate}|\text{Achieve A ratin}$
This allows me to simply estimate the average turnover rate among A lawyers who will eventually be assigned A ratings. The exogenous turnover rate is estimated to be about 7.4%, out of an unconditional turnover rate of 12.4%.

As an indirect test of the model, I also estimated turnover rates separately by year of experience. The model predicts that these experience-conditional turnover rates should be identical for lawyers who will achieve A ratings (and whose turnover is therefore exogenous), while it should decline over time for lawyers who will not receive A ratings. Checking whether this is indeed the case serves to jointly test the assumption that the displacement rate is constant across the career, and the model's deeper assumption that poached lawyers are adversely selected. In what appears to be a resounding success for the model, I am unable to reject the hypothesis that turnover rates are identical by year of experience for non-separating lawyers. By regressing turnover on a constant and year-of-experience dummies, the p-value for the F-statistic is 0.21 for lawyers who eventually achieve A ratings, and below 0.001 for lawyers who eventually don't obtain A ratings. Another way that the estimation of λ^D appears to validate the model is that the turnover rate in the data converges to approximately the estimated value, suggesting (as is predicted by the model) that poaching shuts down once the resumé is sufficiently narrow.

Estimation of ϕ and γ. I estimate ϕ and γ using the result in Proposition 7. Letting $\tilde{v}_i = \tilde{v}(\mathbf{x}_i)$ denote expected log permanent income conditional on $\mathbf{x}_i$ and $p_{A,i}$ denote the probability of A ratings attainment conditional on $\mathbf{x}_i$, Proposition 7 says

$$\tilde{v}_i = \tilde{v}_0 - \phi \ln\left(1 - \tilde{\gamma} p_{A,i}\right).$$

I will use a vector of law school dummy variables as the instrument $\mathbf{x}_i$. I assume that log housing expenditure equals log permanent income, plus idiosyncratic noise u_i that is independent of an individual's law school.[45] This assumption requires that the choice to spend more or less than the average fraction of one's permanent income on housing is exogenous to law school attendance. It is even fine for individuals to have attended their chosen school on the basis of their desire to consume more housing, so long as this choice did not on average translate into spending a larger *share* of income on housing.[46] My

[45]This implicit unit-income elasticity is in line with estimated elasticities for the sample period, which are reviewed in Wilkinson (1973).

[46]For example, individuals could defer in their preference for leisure as opposed to consumption, but behave identically with respect to the share of consumption expenditure allocated to housing.

measure of housing wealth will contain recorded monthly rental payments for renters, and imputed monthly user-costs of housing for home-owners. The user-cost imputation follows the strategy of Albouy and Zabek (2016), who used the same Census dataset.[47] One could imagine using other instruments–the important thing is that the instrument(s) influence the individual's talent prior to labor market entry and are uncorrelated with idiosyncratic housing preferences.

Thus, $\tilde{h}_i = \ln H_i = \tilde{v}_i + u_i$, and

$$\tilde{h}_i = \tilde{v}_0 - \phi \ln\left(1 - \tilde{\gamma} p_{A,i}\right) + u_i. \tag{2}$$

To provide some intuition for the above equation, suppose that $\tilde{\gamma}$ equals one. In this case, ϕ measures the elasticity of housing wealth with respect to the percentile of the school's marginal alum who is expected to get the A rating, $1 - p_A(LSQ_i)$. Suppose that the marginally A rated lawyer would be a median alum from Harvard, but a 90th percentile alum from a relatively unknown local law school. In this case, we can conclude that the average Harvard alumni earns $\frac{0.9}{0.5}\phi = 1.8\phi$ log points more net present income than the average alum from the unknown school.

A challenge in estimating Equation 2 is that it requires knowledge of $p_{A,i}$. Because the equation is non-linear in this term, and $\tilde{\gamma}$ is identified off of curvature, estimation error in $p_{A,i}$ could bias estimates of $\tilde{\gamma}$ that treat these estimates as data. To get around this challenge, I will invert Equation 2 to get $p_{A,i}$ on the left-hand-side.

$$p_{A,i} = \frac{1}{\tilde{\gamma}} \left(1 - e^{\frac{1}{\phi}\left(\tilde{v}_0 - \left(\tilde{h}_i + u_i\right)\right)}\right).$$

Take an expectation over $\mathbf{x}_i$, let $\bar{h} = \mathbb{E}[\tilde{h}_i]$ (the unconditional expectation), and let $k = \mathbb{E}[e^{\frac{1}{\phi}(\tilde{v}_0 - u_i - \bar{h})}|\mathbf{x}_i]$. k is constant by independence of u_i to $\mathbf{x}_i$. This yields

$$p_{A,i} = \frac{1}{\tilde{\gamma}} \left(1 - k e^{-\frac{\tilde{h}_i - \bar{h}}{\phi}}\right). \tag{3}$$

Note that the conditional expectations $\tilde{h}_i$ must be estimated for each law school. Because log housing wealth varies continuously in the population, this term will be much easier to estimate than $p_{A,i}$ would have been. Take a second-order Taylor expansion of Equation 3 with respect to $\tilde{h}_i$ around $\bar{h}$, to get

[47]Specifically, I multiply home values by 0.0789 to get an annual imputed rent, and then divide by 12.

$$p_{A,i} \approx \underbrace{\frac{1-k}{\tilde{\gamma}}}_{b_0} + \underbrace{\frac{k}{\tilde{\gamma}}\phi}_{b_1} \left(\tilde{h}_i - \bar{h}\right) + \underbrace{\frac{k}{2\tilde{\gamma}}\phi^2}_{b_2} \left(\tilde{h}_i - \bar{h}\right)^2. \tag{4}$$

By estimating the above equation via Ordinary Least Squares (OLS), estimates of ϕ and γ can be inferred via $\phi = 2\frac{b_2}{b_1}$ and $\tilde{\gamma} = \frac{1}{b_0 + \frac{b_1^2}{2b_2}}$.

A potential problem with estimating Equation 4 is that different locations may have had different standards for the assignment of A ratings. It is tempting to deal with this possible issue by simply including additive location fixed effects into the equation. However, the theory from which I derived this equation, which is in the proof for Proposition 7, says that changes in the talent threshold required to get an A rating will cause changes in both intercepts and slopes. Thus, in order to account for differences across geographical markets (indexed by m), the equation should become

$$p_{A,i,m} = b_{0,m} + b_{1,m}\left(\tilde{h}_i - \bar{h}_m\right) + b_{2,m}\left(\tilde{h}_i - \bar{h}_m\right)^2. \tag{5}$$

I first estimate the equation pooling across counties. Next, I cut the sample into three market-size categories.[48]

The estimates of ϕ and γ are presented in Table 4. The results are plausible for largest market-size category, which is where most of the law firms are located (I do not limit this sample to lawyers who were only in firms). The estimate of ϕ in the pooled specification seems implausibly large. As anticipated above, the apparent problem here is that lawyers from better law-schools self-select into locations where, in an absolute sense, it is harder to qualify for an A rating. The lawyers from the best schools tend to all be in cities where the standards for A ratings are comparable, so the "good" variation in A ratings attainment is between lawyers with relatively high $\tilde{h}_i$. This biases the curvature in the relationship between A ratings attainment and permanent income by law school. Also, ϕ is estimated very imprecisely for markets with less than 100 lawyers. The lawyers in these small markets tend to have all gone to relatively low quality law schools, so there is insufficient variation in $\tilde{h}_i$ to estimate anything in the small markets. The larger markets (with more than 500 lawyers) are the main focus of the analysis, because they contain most of the law firms. Thus, my preferred estimates are $\hat{\phi} = 2.16$, and $\gamma = 6.03$.[49]

Having estimated γ, ϕ, and λ^D, I will now estimate the final parameter α using the

[48]However, the $\tilde{h}_i$s themselves are estimated in a pooled sample without controlling for geographical markets.

[49]I experimented with splitting the sample by individual city, and I found the results to be qualitatively similar to my preferred estimates, but I lack statistical power to do this for more than a few cities.

simulated method of moments (SMM, see McFadden (1989)). As suggested by the identification proof, I will choose the value that minimizes differences between empirical gross turnover rates and those predicted by the model. Gross turnover will include both poaching and displacement transitions, following the terminology of Section 3. However, because the exogenous displacement rate λ^D was already estimated, the only free margin along which the simulated model can match gross turnover is through poaching.

I will simulate $t = 1, ..., T$ years of labor market experience for $s = 1, ..., S$ individuals, each with initial resumé $[1, \gamma]$,[50] by choosing random values of innate talent $z = z_1, ..., z_S$ and random sequences of displacement shocks $d = \{d_{1,1}, ..., d_{1,T}\}, ..., \{d_{S,1}, ..., d_{S,T}\}$. Fixing these simulated elements, a given guess of α will translate into a deterministic sequence of poaching outcomes. Let $q_{s,t}(\alpha)$ be a binary indicator for whether, in the tth year of labor market experience, the simulated individual experiences turnover. Meanwhile, let i denote an actual individual in the data and let $q_{i,t}$ denote whether or not they experience turnover in year t. The simulated time t turnover rate is

$$\tau_t^S(\alpha) = \frac{1}{S} \sum_{s=1}^{S} \tau_{s,t}.$$

The model error for observation i, t is

$$\varepsilon_{i,t}(\alpha) = q_{i,t} - \tau_t^S(\alpha).$$

The tth moment is $e_t(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t}(\alpha)$. Let $e(\alpha) = \Big(e_1(\alpha), ..., e_T(\alpha)\Big)^T$. Fixing the simulated elements, the SMM estimator of α is

$$\hat{\alpha} = \arg\min_{\alpha} Q(\alpha), Q(\alpha) = e(\alpha)^T W e(\alpha),$$

where W is a positive definite weight matrix. To choose the weight matrix and estimate the asymptotic covariance matrix of the estimates, I use the general indirect inference procedure described in Gourieroux et al. (1993), p.S92. Accordingly, the optimal weight matrix, $W^\star$, is the inverse of the variance-covariance matrix of the moments (evaluated at the true parameter).[51]

Solving the Model. I will now explain how I computationally solve the model, which is necessary to produce the simulated moments above. To solve the model analytically,

[50] All presented results will be scale-invariant to the resumé.

[51] The covariance matrix is estimated via bootstrapped estimation of the entire set of experience-dependent turnover rates. The bootstrap is blocked on individual to account for within-individual clustering of turnover, which is implied by the model.

one would compute the equilibrium objects described in the one-period model, and then apply backward recursion using the propositions in Section 4 to calculate the equilibrium cutoff rule $\zeta_t(\theta, z_1, z_2)$, the ex ante incumbent profit function $\Pi_t^I(\theta, z_1, z_2)$, and the promised value function $V_t(z_1, z_2)$. Because these functions lose analytical tractability after a few iterations, it is essential to use numerical methods.

Thus, I solve for these functions along discrete grid-points and use interpolation when the recursive procedure demands knowledge of an off-grid value. One particularly useful feature of the model is that every equilibrium object features homogeneity. Thus, I only need to explicitly solve the equilibrium functions in the case where $z_2 = 1$. To evaluate functions in cases where $z_2 \neq 1$, and simply apply homogeneity. For example, $V(z_1, z_2) = z_2^{\phi} V(\frac{z_1}{z_2}, 1)$.

The Markov equilibrium can be approximated arbitrarily well by setting the number of periods and grid-points to be sufficiently large. I repeatedly solve the finite horizon model backwards until firm cutoff rules and worker offer-acceptance rules converge to stationary functions.[52]

5.3 Estimation Results

All parameter estimates (or calibrated values) and standard errors are listed in Table 5. The first standard error listed for α does not account for estimation error in the other parameters. To get a conservative upper bound on the correct standard error, I drew from the upper and lower bounds of the 95% confidence intervals for λ^D, ϕ, and γ, and estimated α separately for all 8 possible combinations. I divided the difference between the largest and smallest estimate by 1.96 to get a "rule of thumb" adjusted standard error.[53]

Results from the simulated model. The simulated and empirical experience turnover profiles are plotted in Figure 8. After year 10, the model predicts that the great majority of turnover is exogenous and not uninformative, causing the simulated turnover rate to asymptote to the estimated value of $\lambda^D = 7.4\%$. Indeed, the empirical turnover rate converges to this value, which is impressive given that λ^D was estimated from the signaling content of early-career retention. Simulated turnover rates in year five are somewhat lower than predicted. In general, the model will always tend to predict a convex

[52] Keep in mind that I will allow the discount factor to be relatively small. Thus, the important difference between a finite horizon and an infinite horizon is not that a career lasts forever—it is simply that the worker does not know *when* their career will end.

[53] I opted for this rule of thumb to save significantly on computing time. However, standard errors for all parameters would ideally be computed by bootstrapping the entire procedure from start to finish, blocking on law school (since that is where much of the important variation used to estimate ϕ and γ lies).

experience-turnover profile without added refinements. The refinements that would help match a flatter turnover profile would be adding frictions that either (1) reduce the frequency by which poaching firms can make outside offers, or (2) reduce the rate at which firms acquire private information about their workers.

The theoretical model predicted that workers initially over-place. This means that initial match quality is distorted compared to a setting with the same amount of public information, but *no* private information. In Figure 9, I plot the optimal placement choice of a currently unattached worker as a *ratio* of the ex-ante efficient placement, for different values of the current resumé (by the homogeneity property, what matters for this plot is the ratio of z_1 to z_2). We see that workers with a high degree of initial uncertainty in the resumé tend to over-place. For example, workers entering the labor market with a ratio of $z_2/z_1 = 6.04$, the estimated value of γ, will place at a firm where their expected output is about 73% of what it would have been if they had placed at the ex-ante efficient firm. As the uncertainty in the worker's resumé decreases, the incentive to over-place subsides.

6 Normative Analysis and Counterfactual Simulations

With the estimated model in hand, I use it to answer some normative questions:

- How much less efficient is the market compared to a full information benchmark?
- How much inefficiency can be attributed to the speed at which the market identifies talent, relative to mismatched assignments that are known to be inefficient based on the information that the market currently has?
- What kinds of policies can enhance efficiency?

Benchmarking efficiency. To benchmark the industry's efficiency, it is useful to compare it to two extreme cases: a fully efficient industry where a worker's talents are immediately revealed to all agents, and a myopic industry where learning is shut down. In both cases, the initial assignment is permanent. Ex ante net present output in the two benchmark cases is easy to derive. We simply assign a worker to her optimal firm type based on the available information, and take an ex ante expectation. The periodic output in the two benchmark cases is constant over time. In the simulated industry, output gradually increases over time due to learning. At some point, learning comes to a halt, because mismatch between workers and firms can no longer overcome the adverse selection problem.

Figure 10 plots equilibrium net present output (and earnings) of a representative worker as a fraction of the full-information benchmark. The figure also plots output (as a fraction of full-information) in the one-period version of the model where the worker enters unattached. We see that the equilibrium features a more than 20% efficiency shortfall. However, the one-shot optimum is substantially *more* inefficient. Thus, reallocation serves an important efficiency enhancing role. As z_1/z_2 increases, the one-shot optimum and equilibrium output both increase in efficiency and become closer together. This captures the intuitive fact that, as initial credentials become more informative, average match efficiency increases, and the importance of reallocation diminishes.

The 20% efficiency shortfall is still quite large, so in the remainder of this section I will evaluate the causes and possible policy solutions to this inefficiency. I structurally decompose misallocation into an informational component and a residual non-informational component. To compute the informational component, I take the endogenous stochastic process of the resumé along the equilibrium path, and make it exogenous. I hand control of allocations over to a benevolent social planner who simply assigns the worker to her ex ante optimal firm based on the information available in each period.[54] Thus, at any point in time, the planner's failure to reach full efficiency simply reflects the absence of information. This thought experiment essentially transforms the information structure into a public learning environment a la Farber and Gibbons (1996). I calculate that the social planner in this scenario would achieve 83% efficiency. Thus, the majority of market inefficiency is informational. The main source of inefficiency is the fact that turnover is almost entirely exogenous, and therefore uninformative, for workers with more than ten years of experience.

The social value of pre-job market signaling. I will now perform a policy counterfactual to study the value of using competition in the educational system as a means of producing better information about workers before they go on the job market. My analysis is intended to shed light on recent trends in skilled professions, where internships appear to be supplanting academic competition as an entryway into top jobs. I will explain the internship trend for lawyers, but similar trends are occurring among MBA students and undergraduates going into finance.

Law school grades have typically been very important determinants of post-law school placements in the US, and schools have traditionally ranked their students academically. However, in recent years, grades in years two and three of law school have become far less consequential for placement. This decline appears related to a rising prevalence of the "2L summer associate-ship," which is essentially an internship occuring in the summer after the second year of law school. A prestigious summer associateship is close to a necessary and sufficient condition for getting a prestigious job after law school. Students

[54] By the first welfare theorem, a competitive market would reach the same static optimization outcomes.

apply for these positions in the fall of their second year of law school and thus have only two semesters worth of grades to use in the application. For the remaining four semesters of school, the top students have already been accepted as summer associates and have much lower incentives to compete academically. Thus, second- and third-year grades in law school are not considered to be very informative, at least compared to grades in the first year.

Internships allow firms to acquire private information about workers. In the case of lawyers, this appears to have recently come at the cost of public information, by reducing the signaling quality of second- and third-year grades. Should we consider banning such internships? To assess this within the framework of my model, I will consider a policy that forces the worker to delay her labor market entry for duration D in order to compete academically. The signal generated by academic competition is the same binary cutoff signal that would have been generated in the first year of the labor market. The benefit of academic competition is that, having received her signal, the worker leaves school unattached and unimpeded by the adverse selection problem that would have limited her ability to efficiently match to a new firm. The cost is the D units of foregone productive time (I will need to assume a discount rate for this calculation). The objective is to calculate how small D needs to be in order to make the extended academic competition worthwhile.

I assume that the representative worker's resumé is $[1, \gamma]$ (the results will be scale-invariant to the resumé), and I use the estimated model to compute her equilibrium expected earnings $V_0 = V(1, \gamma)$ and the cutoff that will be used in her first period of tenure at the incumbent firm, $\zeta_1 = \zeta(\theta(1, \gamma), 1, \gamma)$. Thus, the academic competition is assumed to reveal whether the worker is above or below ζ_1. The value of academic competition is

$$V_A = e^{-rD} \left(\frac{\gamma - \zeta_1}{\gamma - 1} \underbrace{V(\zeta_1, \gamma)}_{\text{Value if wins competition}} + \frac{\zeta_1 - 1}{\gamma - 1} \underbrace{V(1, \zeta_1)}_{\text{Value if loses competition}} \right).$$

I find that $\frac{V_A}{V_0} = 1.12$. Assuming an interest rate of $r = 0.05$, this means that the maximum duration D at which academic competition is worthwhile is 2.3 years. Even if the competition required administrative costs, they could exceed five percent of her lifetime earnings and she would still find it worthwhile to participate in a competition lasting one year. These results suggests that, for workers entering the skilled professions, the social returns to education can easily be justified on the basis of signaling via academic competition. The results further suggest that internship trends in law and related professions may be socially inefficient, even if they are privately attractive to the firms that offer them.[55]

7 Conclusion

This paper presented evidence of an inverted job ladder in law, where poached workers move to worse firms while displaced workers move to better firms. I developed a new theory predicting that inverted job ladders will tend to arise in labor markets characterized by private employer learning and comparative advantage in the utilization of talent—two features that are supported by my data on lawyers and which seem present in most of the skilled professions, such as consulting, finance, accounting, marketing, and engineering, where work is team-based and it is thus difficult for an individual member to externally verify how large her contributions were.

The model used here extends the literature on private learning by allowing for infinite horizon Markov dynamics, compared to the two- or three-period models of the past. Allowing for a long time horizon is important for two reasons. First, a longer horizon increases the potential accumulation of strategic inference, which lessens the degree of misallocation. Second, a longer horizon increases the costs of the stigma associated with being rejected by one's firm, which increases the leverage that incumbent firms have over their workers, and increases the degree of misallocation. By accounting for both of these factors, my model more realistically portrays the misallocation costs of private information than the previous generation of private learning models.

By structurally estimating and solving the model, I found that the market for lawyers is less than 80% as efficient as the full-information optimum. This inefficiency is in large part due to a lack of information about talent that *never* becomes resolved, but it is also caused by non-informational distortions in how workers are placed, which would be eliminated in an alternative environment with the same amount of ublic information, but no private information. Thus, the endogenous retention and poaching decisions in the marketplace are highly informative, but adverse selection prevents this information from being immediately reflected in how workers are allocated. My counterfactual analysis suggests that the use of academic competition in the educational system, at the expense of delayed labor market entry, is likely to be efficiency enhancing. The rising importance of internships as a path to full-time employment in the skilled professions, which, at least in the case of law, appears to have come at the expense of academic competition, could

[55]Here I am suggesting the possibility of a coordination failure, in which all students would ex ante be better of by collectively competing in school, but have little incentive to compete given that they expect no one else to.

for the same reason be reducing efficiency.

I view my paper as complementing, rather than challenging, the standard job ladder literature. Although that literature has been extremely successful at explaining labor reallocation in the aggregate economy, the standard job ladder hypothesis appears unlikely to hold specifically in the skilled professions—an important and growing part of the labor market. Consequently, researchers using the job ladder framework in future work ought to consider a separate treatment for the skilled professions.

Future research. This paper opens the door for much future research. One promising avenue is to continue testing the predictions of the inverted job ladder with modern datasets covering a variety of professions. Is there an inverted job ladder in finance, accounting, or consulting? One possible dataset to use for this would be the Survey of Income and Program Participation (SIPP). Because the SIPP was used in Moscarini and Postel-Vinay (2017) to test some of the standard job ladder predictions, it would be particularly interesting to see how the results change when skilled professions are examined separately from the rest of the labor market. Other promising datasets would be worker-firm datasets where, similar to the dataset that I developed here, the researcher can observe the quality of workers' university and/or graduate-level training so as to rank firms based on their propensity to recruit from top schools and to study mobility through the ranks.

A second more ambitious direction for future work is to continue testing and refining the theory of the inverted job ladder and develop it into a fully integrated model that can serve as a private learning alternative to the public learning models following the tradition of Farber and Gibbons (1996). The model makes quantitative predictions about relationships between tenure, turnover, and wage dynamics, which can all be tested in modern panel datasets like the SIPP.

I envision several promising refinements of the inverted job ladder theory. The first refinement is to add promotions, an important channel for strategic information transmission in modern skilled professions such as law, and a central focus of the private employer learning literature. The second refinement would be to add uncertainty over reasons for separation. The current model makes the rather stylized assumption that potential employers can perfectly distinguish exogenous displacements from endogenous separations. To the extent that "rejected" employees can mask their rejection by claiming that their previous employer was a poor fit, the current model may overstate the stigma of separation and thus the leverage of incumbent firms. A third refinement would be to add gradualism to the private learning process, and a fourth refinement would be to allow for a more flexible production function.

Figure 1: Martindale-Hubbell's confidential key (1931 edition)

CONFIDENTIAL KEY

to

The Martindale-Hubbell Law Directory

Numerals immediately following name indicate years of birth and admission.

Estimate of Legal Ability

NOTE—No arbitrary rule for determining legal ability can be formulated. Ratings are based upon the standard of ability for the place where the lawyer practices. Age, practical experience, class of practice, with other necessary qualifications are considered; reports are obtained through various channels and we endeavor to reflect the consensus of reliable opinion.

To qualify for "a", lawyers must be reported "very high" and have been practicing not less than ten years. To qualify for "b", lawyers must be reported "high" and have been practicing not less than five years. A lawyer reported "very high" and in practice more than five years but not long enough to qualify for "a" is rated "b".

"a", very high; "b", high; "c", fair.

Recommendations

NOTE—Nothing derogatory should be inferred from absence of rating. "v", very high.

Estimated Worth

NOTE—It is often difficult to get reliable estimates, therefore the ratings given must be considered as approximations only.

1 estimated over $100,000			5 estimated from	$10,000 to	$20,000
2 " from	50,000 to	$100,000	6 " "	5,000 to	10,000
3 " "	30,000 to	50,000	7 " "	2,000 to	5,000
4 " "	20,000 to	30,000	8 " "	1,000 to	2,000

9, estimated less than $1,000

Rating for Promptness in Paying Bills

"g", good; "f", fair; "m", medium.

An asterisk (*) following the name of place indicates a County Seat.

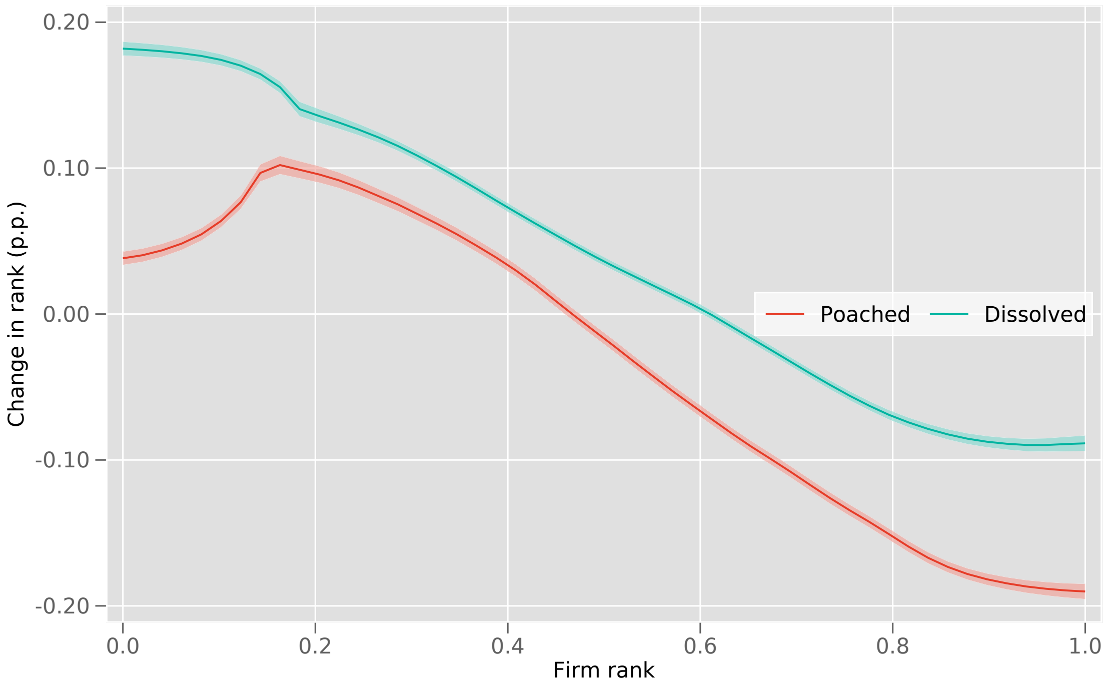
"§" does not want collections.

"‡" not in general practice. This includes those engaged chiefly in other occupations than that of law, those retired, and those not in active practice.

"©" This character indicates that the lawyer after whose name it appears is listed at more than one point.

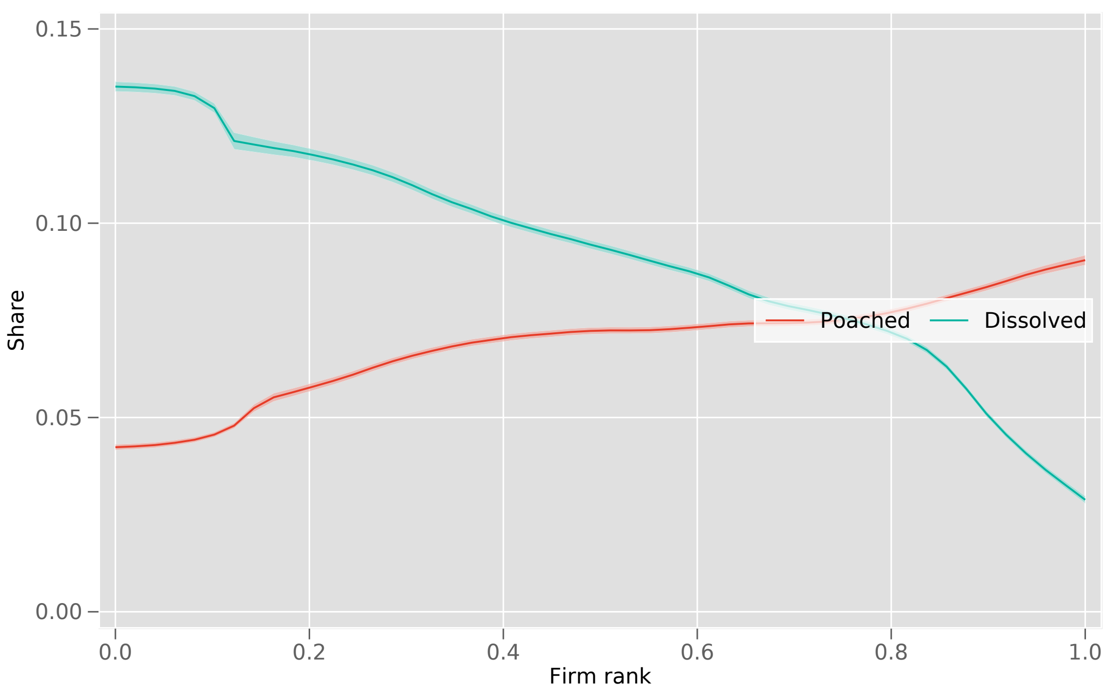
NOTE—Absence of rating characters (whether indicated by blank space or dash "—") should not in any case be construed as derogatory to anyone. This may mean that sufficient information was not obtainable up to time of going to press. Also, in some places we do not publish complete ratings or rate all who are worthy.

Figure 2: Dynamics of firm rank, conditional on current rank



Kernel density estimates with 95% CI

Figure 3: Rates of poaching and dissolution



Kernel density estimates with 95% CI

Figure 4: Future A ratings attainment by current tenure, $r_{i,t} > 0.75$

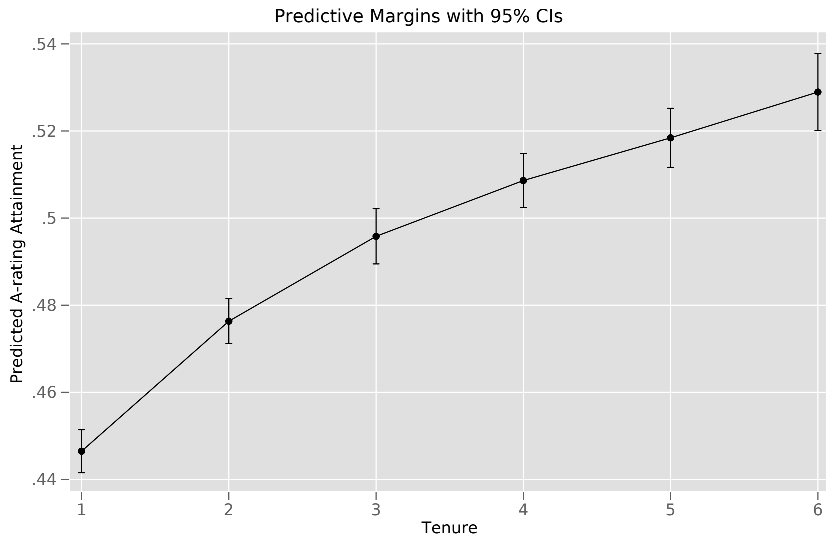


Figure 5: Poaching rates by current tenure, $r_{i,t} > 0.75$

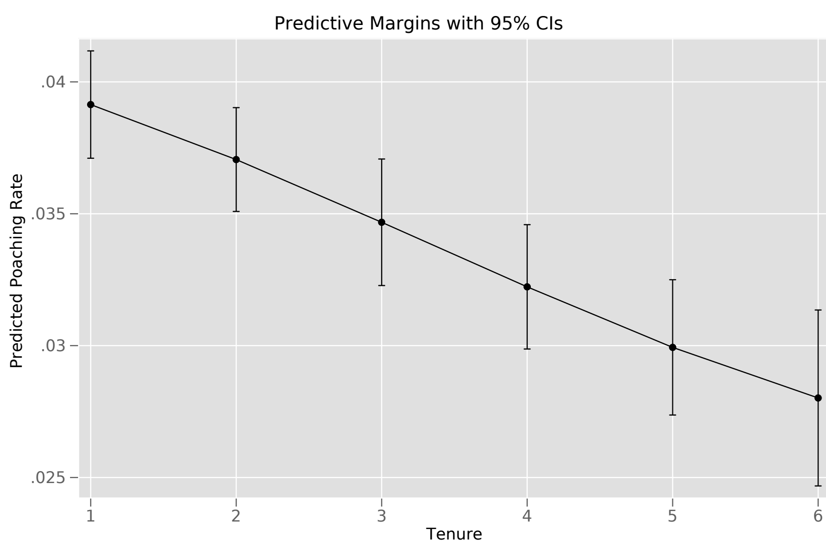


Figure 6: Graphical rendition of two period game

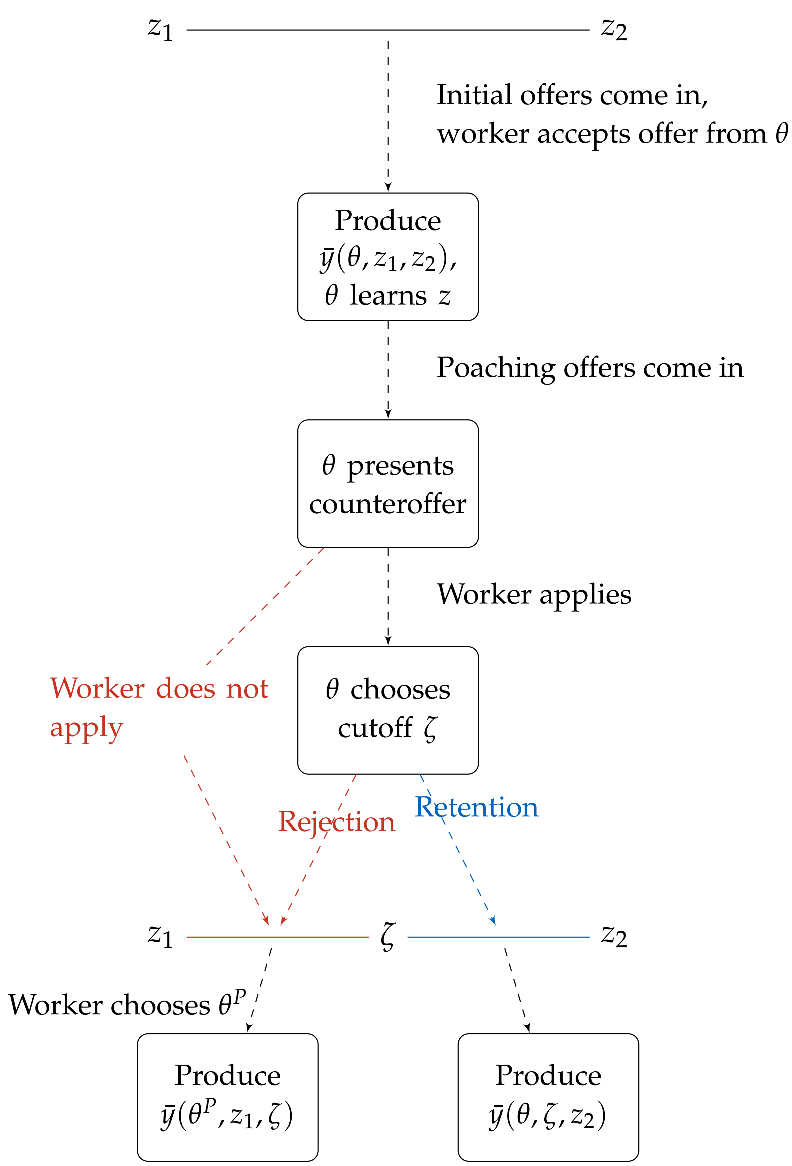


Figure 7: Graphical rendition of infinite horizon game

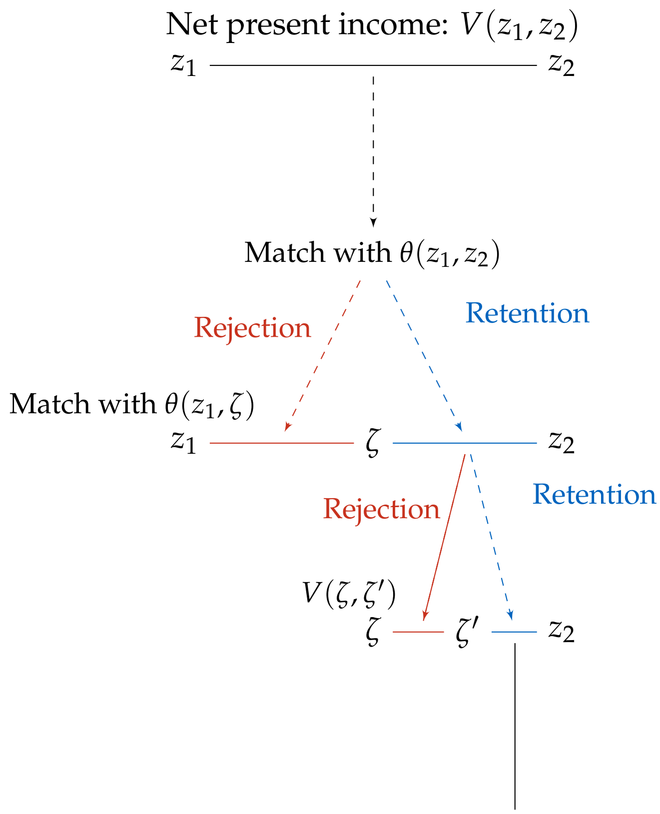


Figure 8: Simulated versus actual turnover

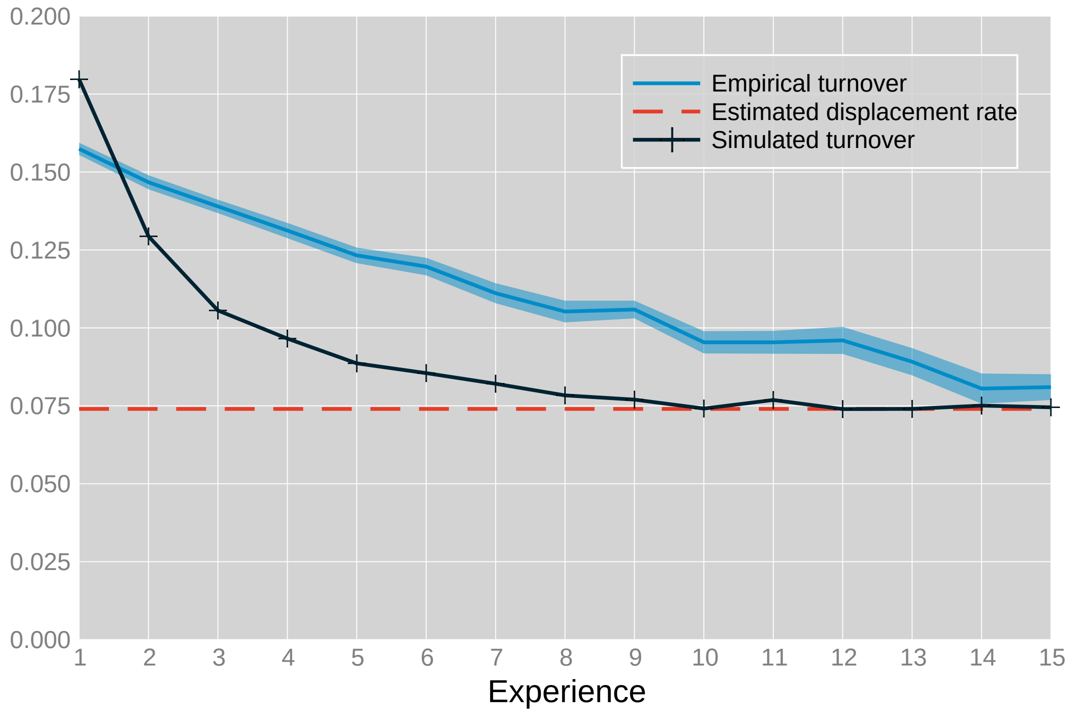


Figure 9: Distortions in equilibrium placement

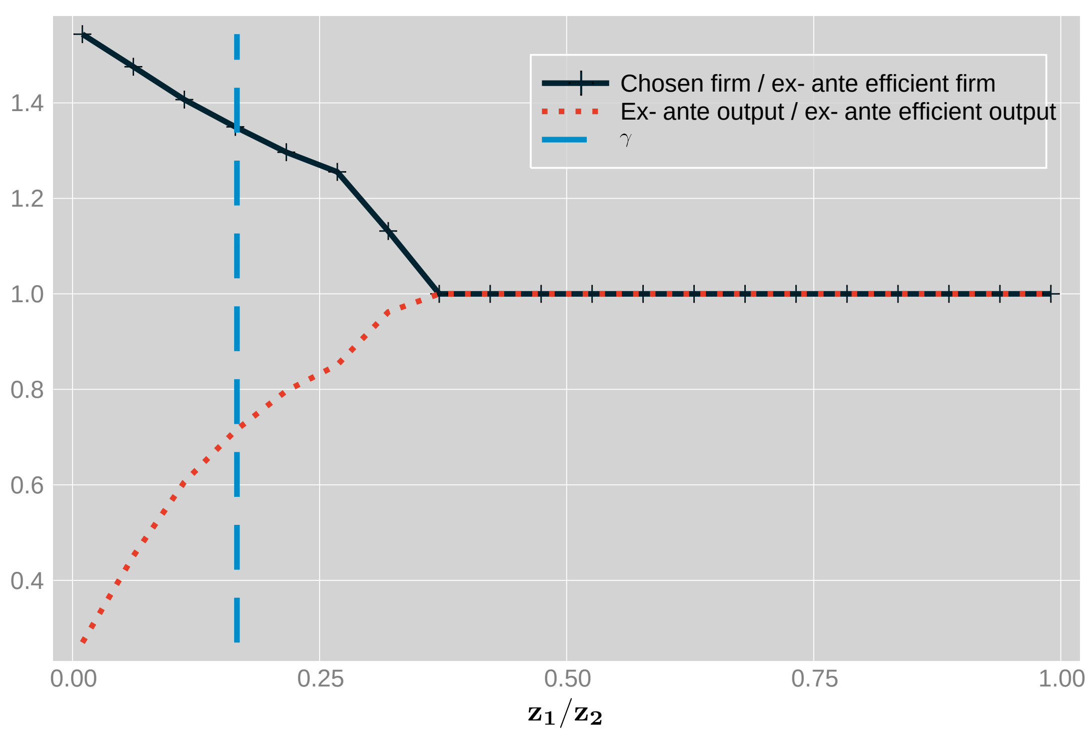


Figure 10: Expected output (% of full-information) of currently unattached worker

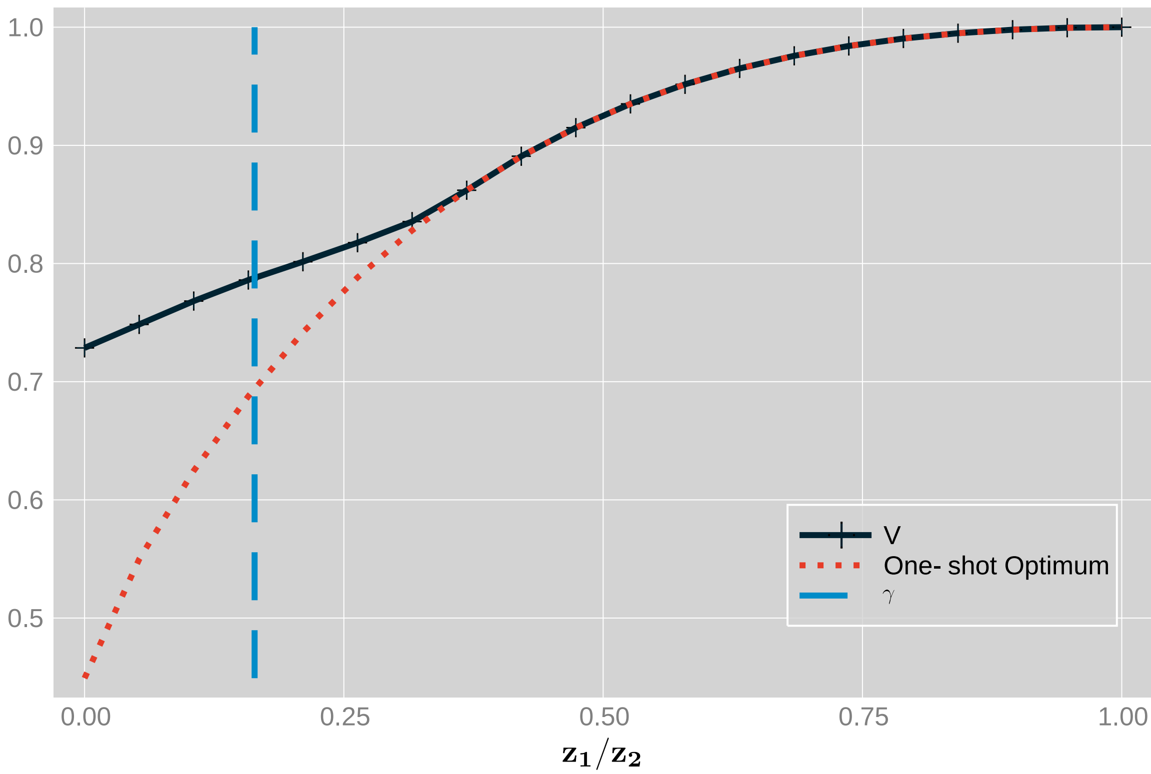


Table 1: Summary statistics by lawyer-year

	Mean	Std.dev.	p.05	p.95
Age	41.06	7.81	29	54
Exper	10.63	7.23	1	24
A rated	0.26	0.44	0	1
Mkt. size	5519.20	7606.29	31	23231
Firm size	16.15	50.96	4	44
	Classified transitions			
Attrit.	0.03	0.17		
Poached	0.05	0.23		
Displaced	0.03	0.17		
Retained	0.83	0.38		
Exit	0.08	0.29		
Obs.	539,342			

Sample: Lawyers currently in law firms of three or more lawyers
Mkt. size reflects number of lawyers working in local town or city

Table 2: Own Law School Quality versus Colleagues' Characteristics

Dependant variable	**LSQ**	
	All obs	New hires only
Avg LSQ	0.566***	0.591***
	(0.001)	(0.005)
Share A-rated	0.080***	0.086***
	(0.001)	(0.006)
Log firm size	0.050***	0.050***
	(0.001)	(0.002)
Log mkt. size	0.002***	-0.005***
	(0.000)	(0.001)
N	1,512,679	88,825
R^2	0.245	0.261

Controls for mkt. size and current age
Sample of age 40+ lawyers matched to 1940 Census
Std. errors clustered on law school (in parens)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Dynamics of firm rank by type of transition

Change in...	**Job leavers**		**Job stayers**
	Poached	Displaced	Retained
	Mean/SE	Mean/SE	Mean/SE
Mean colleague LSQ (z-score)	-0.035	0.027	-0.001
	(0.002)	(0.002)	(0.000)
Firm size	-6.933	6.127	0.146
	(0.079)	(0.055)	(0.002)
Share colleagues A-rated	-0.024	-0.020	0.003
	(0.002)	(0.002)	(0.000)
Firm rank (p.p.)	-0.058	0.046	-0.001
	(0.001)	(0.001)	(0.000)
Share of total	0.0274	0.0912	0.8189
Obs.	40,112	133,726	1,200,682

Sample: Lawyers remaining in private-practice law

Table 4: Estimation of ϕ and γ

	Estimate (std. err.)		Observations
Sample	ϕ	γ	
Split by Market Size Categories			
>500 lawyers	2.16 (0.38)	6.03 (0.35)	74,212
$100-500$ lawyers	2.08 (1.63)	5.24 (0.627)	20,413
<100 lawyers	3.67 (6.60)	5.59 (0.263)	38,721
Pooled			
	7.29 (1.91)	6.77 (0.09)	133,346

Controls for mkt. size and current age
Sample of lawyers in 15th year
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors

Table 5: Model Estimates

Parameter	**Estimate** (std. error)	**Description**
δ	0.920	1 - Exit rate
λ^D	0.074 (0.002)	Displacement rate
ϕ	2.165 (0.383)	Gross return to talent in production
γ	6.031 (0.352)	Initial value of z_2/z_1
α	1.86 (0.010) | (0.1014‡)	Complementarity in production

Standard errors computed via bootstrap (blocked on individual for λ^D and law school for ϕ and γ).
‡ *Adj. standard error using rule of thumb to account for first-stage estimation error.

Appendices

Appendix A Evidence of adverse selection for bottom three firm-rank quartiles

Figure 11: Future A ratings attainment by current tenure, $0.5 < r_{i,t} \leq 0.75$

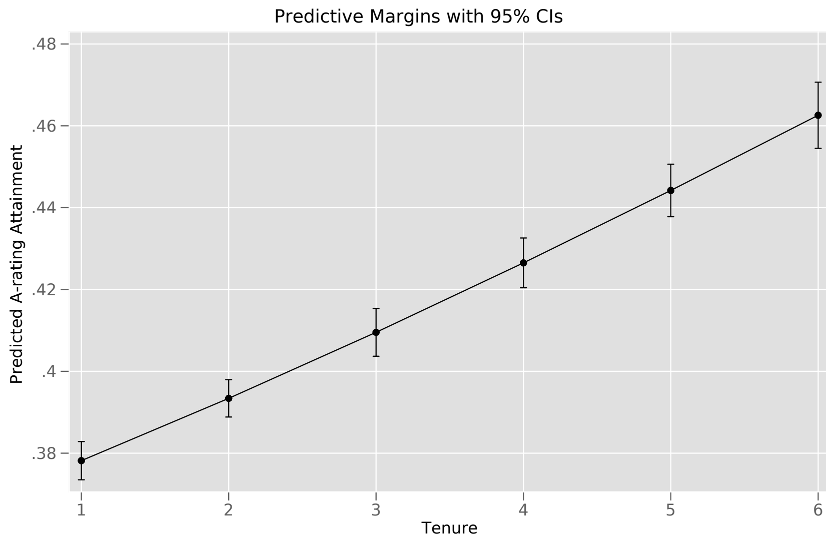


Figure 12: Poaching rates by current tenure, $0.5 < r_{i,t} \leq 0.75$

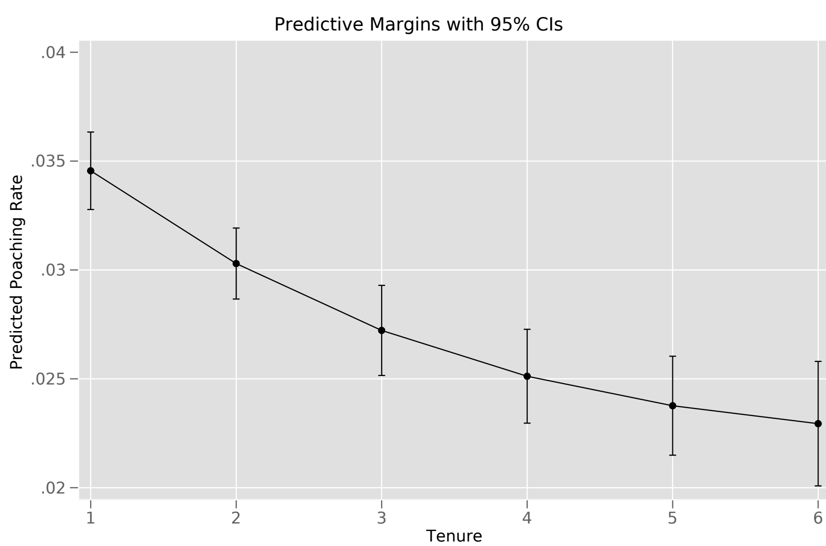


Figure 13: Future A ratings attainment by current tenure, $0.25 < r_{i,t} \leq 0.5$

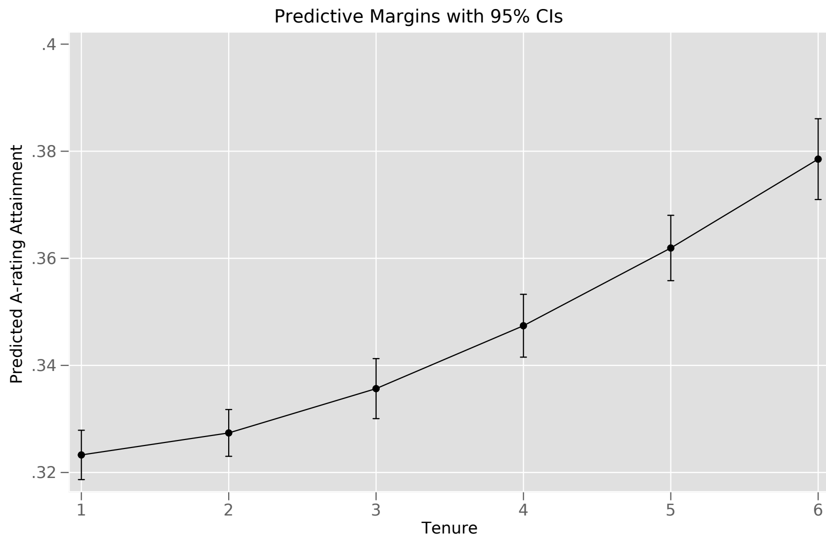


Figure 14: Poaching rates by current tenure, $0.25 < r_{i,t} \leq 0.5$

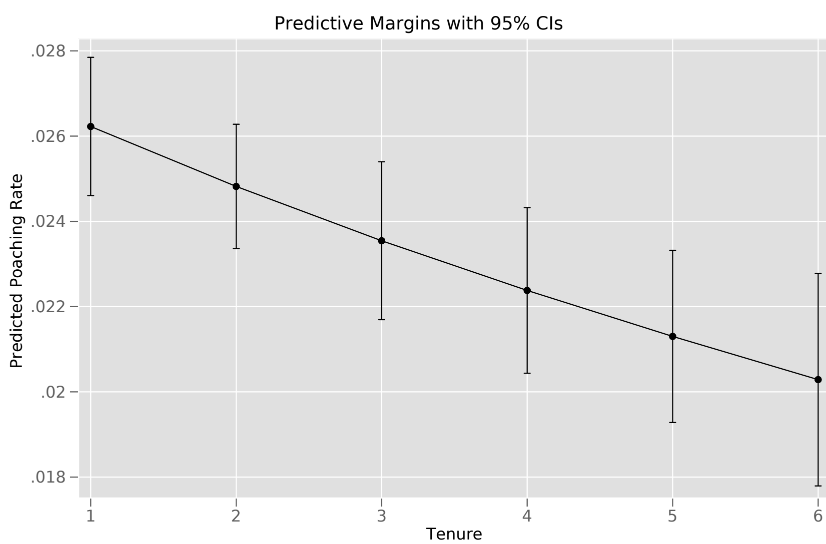


Figure 15: Future A ratings attainment by current tenure, $r_{i,t} \leq 0.25$

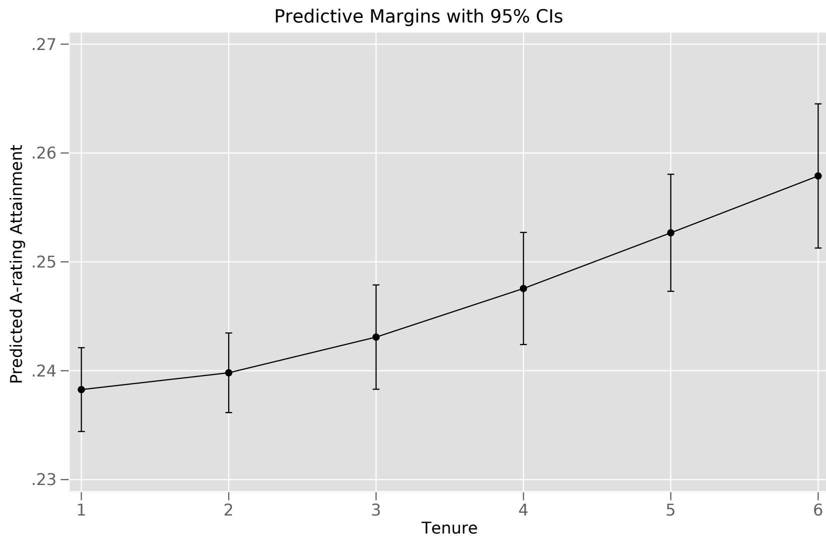
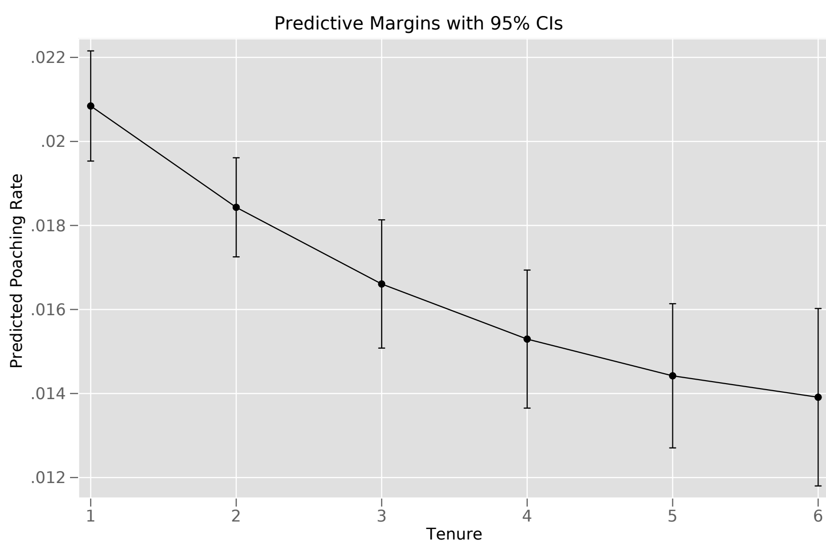


Figure 16: Poaching rates by current tenure, $r_{i,t} \leq 0.25$



Appendix B Analytical derivations of the objects in Definition 1

Using the assumed functional form in Assumption 7, here I analytically derive optimal output under various degrees of information about the worker's talent, defined in **??**.

The full information output is

$$\begin{aligned}\bar{y}^{FIM}[z_1, z_2] &= \mathbb{E}_{z|z_1 \leq z \leq z_2} y(z, z) = (\phi - \alpha)\mathbb{E}_{z|z_1 \leq z \leq z_2}\left[z^{\alpha+(\phi-\alpha)}\right] \\ &= (\phi - \alpha)\frac{z_2^{\phi+1} - z_1^{\phi+1}}{(z_2 - z_1)\,(1 + \phi)}.\end{aligned} \tag{6}$$

Again, the last line imposes the uniform distributional assumption. The shortfall of optimal expected output under incomplete information, as compared to expected output under full information, is increasing in the difference $z_2 - z_1$.

$$\begin{aligned}\theta^\star[z_1, z_2] &= \arg\max_\theta \mathbb{E}_{z|z_1 \leq z \leq z_2} y(\theta, z) \\ &= \left(\mathbb{E}_{z|z_1 \leq z \leq z_2} z^{\phi-\alpha}\right)^{\frac{1}{\phi-\alpha}} \\ &= \left(\frac{z_2^{1+\phi-\alpha} - z_1^{1+\phi-\alpha}}{(z_2 - z_1)(1 + \phi - \alpha)}\right)^{\frac{1}{\phi-\alpha}},\end{aligned}$$

where the last line imposes the uniform distributional assumption. The ex ante optimal output is

$$\begin{aligned}\mathbb{E}_{z|z_1 \leq z \leq z_2} y(\theta^\star, z) &= (\phi - \alpha)\left(\mathbb{E}_{z|z_1 \leq z \leq z_2} z^{(\phi-\alpha)}\right)^{\frac{\phi}{\phi-\alpha}} \\ &= (\phi - \alpha)\left(\frac{z_2^{1+(\phi-\alpha)} - z_1^{1+(\phi-\alpha)}}{(z_2 - z_1)(1 + (\phi - \alpha))}\right)^{\frac{\phi}{\phi-\alpha}}.\end{aligned} \tag{7}$$

Appendix C Omitted Proofs

Lemma 1 (The time-T cutoff)**.** Recall that $\bar{y}^{\max}(z_{1,t}, z_{2,t})$ is the maximized expected output of a worker with resumé $[z_{1,t}, z_{2,t}]$, and ζ_T is the equilibrium cutoff rule. The equilibrium cutoff rule is the largest ζ_T satisfiying

$$y(\theta, \zeta_T) \leq \bar{y}^{\max}(z_{1,T}, \zeta_T) \tag{1}$$

The equilibrium cutoff, $\zeta_T(\theta, z_1, z_2)$ is the maximum value within $[z_{1,t}, z_{2,t}]$ satisfying Equation 1.

Proof of Lemma 1. $\bar{y}^{\max}(z_1, \zeta)$ will be the value of the poaching wage when poaching firms correctly anticipate that the cutoff rule ζ is going to be used. Given that it matches this poaching wage, the marginally retained worker produces $y(\theta, \zeta)$. Thus, as long as $\zeta \in (z_1, z_2)$, the incumbent will be indifferent to retaining the marginally retained worker, and thus $y(\theta, \zeta) = \bar{y}^{\max}(z_1, \zeta)$, $y(\theta, z) > \bar{y}^{\max}(z_1, \zeta) \forall z > \zeta$, and $y(\theta, z) < \bar{y}^{\max}(z_1, \zeta) \forall z < \zeta$. It may also be possible to have $\zeta = z_1$, in which case the incumbent may strictly prefer to retain all worker types, or $\zeta = z_2$, in which case the incumbent strictly prefers to reject all worker types.

Existence of at least one value of ζ satisfying $y(\theta, \zeta) - \bar{y}^{\max}(z_1, \zeta) < 0 \forall z < \zeta$ is trivial. Because this function is continuous, it either lies uniformly above 0 (so $\zeta = z_2$ works), uniformly below 0 (so $\zeta = z_1$ works), or crosses 0 at some point (by the intermediate value theorem).

To understand why the equilibrium cutoff must be the supremum of all cutoffs, ζ, ensuring that the incumbent is indifferent to retaining the marginal worker type, we need to think about the incentives of the poaching firms. There is always some poaching firm that can trigger ζ to be used by offering $\bar{y}^{\max}(z_1, \zeta)$—in this case, the poaching firm would be $\theta^\star(z_1, \zeta)$.

Now suppose that in equilibrium, this pivotal offer were not being made. I will prove that there would then have to be some firm slightly below $\theta^\star(z_1, \zeta)$ that could generate strictly positive profits by triggering a cutoff slightly below ζ. Since ζ is, by assumption, the largest value satisfying Equation 1, it must be the case that either $\zeta_t = z_2$ and $y(\theta_t, \zeta_t) - \bar{y}^{\max}(z_{1,t}, \zeta_t) > 0$, or that $y(\theta_t, \zeta_t) - \bar{y}^{\max}(z_{1,t}, \zeta_t)$ is equal to 0 and is increasing in ζ_t (otherwise ζ_t could not be the largest value satisfying Equation 1). If we are in the first case, this means that a poaching firm could profitably offer strictly less than $\bar{y}^{\max}(z_{1,t}, \zeta_t)$ and still induce the incumbent to use the same cutoff (since no one else was already offering a pivotal wage this high). In the second case, this means that a poaching firm could profitably offer $\bar{y}^{\max}(z_{1,t}, \tilde{\zeta}_t)$, for some $\tilde{\zeta}_t$ that was slightly below ζ_t, and cause the incumbent to use a cutoff slightly below ζ_t. We know this would be profitable because $y(\theta_t, \zeta_t) - \bar{y}^{\max}(z_{1,t}, \zeta_t)$ is increasing in ζ_t. □

Lemma 2. Suppose that $z_{1,T} \leq \zeta_T(\theta_T, z_{1,T}, z_{2,T}) < \theta_T < z_{2,T}$. Then the equilibrium poaching firm is of lower type than the incumbent firm.

Proof of Lemma 2. The worker's talent is below ζ with probability one. Thus, the ex post optimal match, $\theta^\star(z)$, is below ζ with probability 1. Because output is concave in firm type, the ex ante optimal firm, $\theta^{\max}(z_1, \zeta)$, must be the ex post optimal firm for some $z \in [z_1, \zeta]$. If θ were higher or lower than this, then output could be improved with probability 1 by decreasing or increasing θ. □

Lemma 3. Conditional on the resumé, attachment status, and incumbent firm, the worker's own beliefs do not affect her payoffs, and it is a dominant strategy to apply for retention.

Proof of Lemma 3. Let the market's beliefs at any time and stage of the game be described by the resumé $[z_{1,t}, \zeta_t]$, and now allow this to differ from the worker's beliefs. The proof will be inductive. First, suppose we are in period $T-1$. In this case, the set of potential poaching offers available to the worker are completely independent of her beliefs. Her best strategy is to take the highest offer, and thus her own beliefs have no impact on her equilibrium earnings in the $T-1$ period subgame.

Now, suppose that we know that in all subgames with t or fewer remaining periods, the worker's payoff will be unaffected by her own beliefs, given her resumé, attachment status, and incumbent firm. Because future payoffs are not influenced by beliefs, the only way that beliefs can affect *current* expected payoffs is by causing a change in the joint distribution of next period's incumbent firm, resumé, and attachment status. This cannot occur if the worker is currently unattached because her resumé and the set of poaching offers will remain unchanged next period. Thus, the only possibility is for a currently attached worker to realize a different distribution of future resumés and attachment status in the next period.

Regardless of what her current beliefs are, there are only three possible outcomes: be displaced and obtain resumé $[z_{1,t}, z_{2,t}]$, remain at the incumbent firm, θ_t, and obtain a resumé $[\zeta_t, z_{2,t}]$, or leave the incumbent firm (without displacement) and obtain resumé $[z_{1,t}, \zeta_t]$. Displacement is exogenous, so the worker's private information can only serve to assign more or less weight to the latter two outcomes.

And now we come to the crux of the proof. Both outcomes deliver the exact same payoff to the worker. Thus, even though the probabilities of the two outcomes can change depending on the worker's private information, the worker's expected payoff cannot.[56]

Given that the worker's payoff is unaffected by her own information, it is a weakly dominant strategy for her to always apply. If there is a positive probability of retention, conditional on applying, then the firm must anticipate strictly positive profit with probability 1 (the knife-edge case is when the worker's talent equals the cutoff, which happens with probability zero). Thus, with probability 1, the incumbent firm would always be willing to pay ϵ more than the worker's otuside option in order to convince her to apply, and thus the worker must always apply in equilibrium. □

Lemma 4.

$$w_t^R(\theta, z_1, z_2, \zeta) = V(z_1, \zeta) - \delta\left((1-\lambda^D)V(\zeta, \zeta(\theta, \zeta, z_2)) + \lambda^D V(\zeta, z_2)\right).$$

Proof of Lemma 4. If the worker is accepted, next period with probability λ^D she will remain attached to firm θ with resumé $[\zeta, z_2]$. The incumbent will then ensure that regardless of whether she is retained or not, she will receive value $V_{t+1}(\zeta, \zeta(\theta, \zeta, z_2))$. Meanwhile, if she is displaced, her resumé will remain at $[\zeta, z_2]$ and she will earn value $V(\zeta, z_2)$. □

Proposition 1. The *equilibrium cutoff rule* is

$$\zeta_t(\theta, z_1, z_2) = \sup_{\zeta \in [z_1, z_2]} \left\{ y(\theta, \zeta) - w_t^R(\theta, z_1, z_2, \zeta) > 0, \forall z > \zeta \right\}.$$

Proof of Proposition 1. □

Proposition 5. For every possible (z_1, z_2, θ), there is some $\bar{\delta}$ such that, if and only if $\delta > \bar{\delta}$, the incumbent's cutoff rule is z_1. $\bar{\delta}$ is decreasing in $\frac{z_1}{z_2}$.

Proof of Proposition 5. Consider the cutoff inequality:

$$\zeta(\theta, z_1, z_2) = \sup_{\zeta \in [z_1, z_2]} \{G(z) < 0, \forall z < \zeta, z \in [z_1, z_2]\},$$

[56]Notice that because of Assumption 3, even if the worker knew she was more talented than $z_{2,t}$, she would never be able to prove this to the market. The market would at best assume that she is of talent $z_{2,t}$.

where

$$G(\zeta;\theta,z_2) = y(\theta,\zeta) + \delta\left((1-\lambda^D)V^I(\zeta,z_2) + \lambda^D V(z_1,z_2)\right) - V(z_1,\zeta).$$

As we make δ arbitrarily high, the expression becomes dominated by the difference in continuation values. The continuation value at the incumbent firm is strictly higher. Thus, there exists some critical value of the discount factor, $\bar{\delta}$, which guarantees that $G(\zeta;\theta,z_2)$ will be uniformly greater than 0 for all $\zeta > z_1$.

As we decrease z_1, we simply expand the interval over which we are considering $G(\zeta;\theta,z_2)$, so the critical $\bar{\delta}$ must be weakly higher than before. □

Corollary 1. There exists a range of values for δ for which there is strictly positive poach when $\frac{z_1}{z_2}$ is sufficiently small, but where $\zeta(\theta,z_1,z_2) < \theta$ at all points along the equilibrium path.

Lemma 5.

$$\frac{\partial^2 \ln y(\theta,z)}{\partial \ln \theta^2}\,|_{\theta=z} = -\alpha\phi.$$

Proposition 6 (Identification of λ^D)**.** Consider a lawyer with a given time t history. Let λ^D denote the rate of exogenous displacement. Assume that the probability of achieving an A rating is strictly between 0 and 1 in the event that the lawyer is retained. Then λ^D is identified by

$$\lambda^D = Pr(\text{Separate}|\text{Achieve A rating}).$$

Proof of Proposition 6. For this proof, let τ_t denote the probability of separation, $p^r_{A,t}$ the probability of th lawyer eventually obtaining an A rating conditional on being retained, and $p^q_{A,t}$ the probability conditional on separating.

Let the incumbent firm's cutoff rule be ζ_t. The turnover rate must satisfy

$$\tau_t = \lambda^D + (1-\lambda^D)\frac{\zeta_t - z_{1,t}}{z_{2,t} - z_{1,t}}.$$

Solving for ζ_t, we find

$$\zeta_t = z_{1,t}\frac{(\tau_t - \lambda^D)\gamma + 1 - \tau_t}{1-\lambda^D}.$$

By assumption, $p_{A,t} \in (0,1)$. This implies that the threshold for obtaining an A rating, z^A, must be in the interior of $[z_{1,t}, z_{2,t}]$. By the additional assumption that $p^r_{A,t} < 1$, we can infer that the incumbent's cutoff rule ζ_t was below z^A. Keep in mind that this implies that any worker who was truly poached should never receive an A rating!

The probability of getting an A rating prior to retention is $p_{A,t} = \frac{z_{2,t}-z^A}{z_{2,t}-z_{1,t}}$. The probability conditional on having been retained—and thus being revealed above ζ_t, is $p^r_{A,t} = \frac{z_{2,t}-z^A}{z_{2,t}-\zeta_t}$. The ratio of the two probabilities is

$$\frac{p^r_{A,t}}{p_A} = \frac{z_{2,t}-z_{1,t}}{z_{2,t}-\zeta_t} = \frac{z_{2,t}-z_{1,t}}{z_{2,t}-z_{1,t}\left(\frac{(\tau_t-\lambda^D)\gamma+1-\tau_t}{1-\lambda^D}\right)} = \frac{\gamma-1}{\gamma-\left(\frac{(\tau_t-\lambda^D)\gamma+1-\tau_t}{1-\lambda^D}\right)}$$

$$= \frac{\gamma-1}{\frac{\gamma(1-\lambda^D)-(\tau_t-\lambda^D)\gamma+1-\tau_t}{1-\lambda^D}} = \frac{\gamma-1}{\frac{\gamma(1-\tau_t)+1-\tau_t}{1-\lambda^D}} = \frac{1-\lambda^D}{1-\tau_t},$$

which implies

$$\lambda^D = 1-(1-\tau_t)\frac{p^r_{A,t}}{p_{A,t}} = \frac{\tau_t P^q_{A,t}}{P_{A,t}} = Pr(\text{Separate}|\text{Achieve A rating}).$$

□

Proposition 7 (Identification of ϕ and γ)**.** Let $\mathbf{x}$ denote the instrument (such as LSQ) described in Assumption 5. Suppose there exists an interval of values for $\mathbf{x}$ where the probability of A ratings attainment conditional on $\mathbf{x}$ is strictly between 0 and 1. Let $\tilde{v}(\mathbf{x}) = \mathbb{E}[\ln V|\mathbf{x}]$, and let $p_A(\mathbf{x})$ denote the probability of becoming A rated conditional on $\mathbf{x}$. Let $\tilde{\gamma} = \frac{\gamma-1}{\gamma}$. Because $\tilde{v}(x)$ and $p_A(\mathbf{x})$ are both increasing, there exists a one-to-one relationship $\tilde{v}(p_A)$. Moreover, ϕ and γ are identified by

$$\phi = \left(\frac{\partial\tilde{v}}{\partial p_A}\right)^2 / \frac{\partial^2\tilde{v}}{\partial p_A^2},$$

and

$$\tilde{\gamma} = \frac{1}{\frac{\partial\tilde{v}}{\partial p_A} / \frac{\partial^2\tilde{v}}{\partial p_A^2} + p_A}$$

Proof of Proposition 7. Consider net present earnings for a new worker, $V = V(z_1, z_2)$. By the previous homogeneity results, we know that $V = z_2^{\phi} V(\frac{z_1}{z_2}, 1) = z_2^{\phi} V(\gamma, 1)$, where the initial value of z_2 is a random variable. Taking logs, letting $k_1 = \ln V(\gamma, 1)$, and applying Assumption 5, we have

$$\ln V = k_1 + \phi \ln z_2 = k + \phi\left(g(\mathbf{x}) + \epsilon\right).$$

Taking expectations yields

$$\tilde{v}(\mathbf{x}) = \tilde{v}_0 + \phi g(\mathbf{x}). \tag{8}$$

Now I will show how to map $\mathbf{E}[\ln z_2|\mathbf{x}]$ into the $p_A(\mathbf{x})$. Assuming that a lawyer obtains an A rating if and only if her talent is above some constant threshold Z_A, then according to Assumption 5, and assuming that z_A is inside of $[z_1, z_2]$ with probability 1, the probability of receiving an A rating conditional on $\mathbf{x}$ is given by

$$\begin{aligned} p_A(\mathbf{x}) = Pr(z > z_A|\mathbf{x}) = \mathbb{E}_{z_1,z_2}\left[\frac{z_2 - z_A}{z_2 - z_1}|\mathbf{x}\right] &= \mathbb{E}_{z_1,z_2}\left[\frac{\gamma z_2 - \gamma z_A}{(\gamma-1)z_2}|\mathbf{x}\right] \\ &= \mathbb{E}_{z_1,z_2}\left[\frac{\gamma \exp\left(g(\mathbf{x}) + \epsilon\right) - \gamma z_A}{(\gamma - 1)\exp\left(g(\mathbf{x}_j) + \epsilon\right)}|\mathbf{x}\right] \\ &= \frac{\gamma}{\gamma - 1}\frac{e^{g(\mathbf{x})} - z_A \mathbb{E}_{z_1,z_2}\left[e^{-\epsilon}|\mathbf{x}\right]}{e^{g(\mathbf{x})}} \\ &= \frac{\gamma}{\gamma-1}\left(1 - k_2 e^{-g(\mathbf{x})}\right), \end{aligned} \tag{9}$$

where the constant k_2 equals $z_A \mathbb{E}\left[e^{-\epsilon}\right]$. Solving for $g(\mathbf{x})$, we have

$$g(\mathbf{x}) = \ln k_2 - \ln\left(1 - \frac{\gamma-1}{\gamma}p_A(\mathbf{x})\right).$$

Plugging this in to Equation 10, and letting the constant $\tilde{v}_0$ equal $k_1 + \phi \ln k_2$, we have

$$\tilde{v}(\mathbf{x}) = \tilde{v}_0 - \phi \ln\left(1 - \frac{\gamma-1}{\gamma}p_A(\mathbf{x})\right). \tag{10}$$

Let $\tilde{\gamma} = \frac{\gamma-1}{\gamma}$. Taking derivatives of $\tilde{v}(\mathbf{x})$ yields

$$\frac{\partial \tilde{v}}{\partial p_A} = \phi \frac{\tilde{\gamma}}{1 - \tilde{\gamma} p_A(\mathbf{x})},$$

and

$$\frac{\partial^2 \tilde{v}}{\partial p_A^2} = \phi \frac{\tilde{\gamma}^2}{(1 - \tilde{\gamma} p_A(\mathbf{x}))^2}.$$

It is now clear how the algebraic manipulations in Equation 10 identify ϕ and γ.

□

Proof of Proposition 8. First, take the current incumbent firm and resumé of a first-year worker exogenous. In this case, the poaching rate is clearly increasing in α, because the mismatch between the marginally retained worker and the incumbent firm becomes relatively more expensive. Given the increased poaching rate, the worker's resumé, conditional on being retained, will improve by a higher margin than before, widening the difference between the incumbent firm, and the firm that she would optimally choose to match with if unattached with the same resumé.

Thus, if the worker's initial placement stayed the same, she would become relatively more under-placed in all future states of the world where she was still at her initial employer. Hence, the increase in α favors higher initial placement, which also causes turnover to increase. $\square$

Appendix D Scoring schools and ranking firms

In order to study the mobility of workers through the ranks, I must rank firms. I rank firms according to a simple principal: higher-ranking firms hire better-credentialed lawyers, and better-credentialed lawyers tend to have gone to better law schools. Thus, enodogenous sorting patterns reveal exogenous technological differences. My procedure has two steps: (1) construct a cardinal measure of law school quality, and (2) rank individuals based on the law school quality of their colleagues. A Harvard graduate surrounded by alumni from no-name schools is assumed to probably be at a low-ranking firm, and a no-name alum surrounded by Harvard graduates is assumed to be at a high-ranking firm. Thus, graduates of bad schools can work at good firms, but they are assumed to be the exception rather than the rule.

Things that I do not control for, but could, include the individual's entire career history and legal ability ratings. Things that I cannot control for include an individual's performance in law school and public case outcomes. These things are no doubt important—in fact, a somewhat obscure Wisconsin Survey of lawyers conducted in 1932 matched the tax returns to the within-cohort academic ranks of 600 graduates of the University of Wisconsin Law School (graduating in years 1914-1932). The study found that higher academic rank was highly predictive of eventual income.[57]

Scoring schools. The first step of the procedure scores law schools based on two measures of the success of their alumni. I use two cardinal outcomes. The first measure is the share of alumni obtaining the (highest possible) *A* MH rating. The second measure is the average alum's MH net worth estimate.[58]

For both A ratings and rent, I need to adjust for differences in location and age. More populated areas are more competitive for ratings, have higher priced real estate, and could disproportionately attract certain law school alumni. Older individuals have had a longer time to build the credentials required for an A rating, may have different demand for housing based on family structure, and may come disproportionately from older law schools. Thus, the A ratings and rent-based measures are constructed as law school fixed effects in a statistical decomposition of each outcome after controlling for a polynomial in age and market size. Since these observations all come from 1940, there is no need to account for temporal differences. For net worth, I need to adjust for secular increases in

[57]See Lloyd K. Garrison (1938), pages 55-56.

[58]I also considered using expenditures on rent and housing using 1940 Census data. Average expenditure was mostly proportional to average net worth. In cases when it was not, it appeared likely to be driven by certain law schools disproportionately feeding into more or less expensive housing markets.

incomes across the sample period, and for the fact that older individuals have had more time to accumulate wealth.

Thus, I statistically decompose each outcome into a law school fixed effect after controlling for the aforementioned factors. To control for secular trends, I include a quadratic polynomial in calendar year. To control for market size, I include a quadratic polynomial in the log number of locally practicing lawyers. To control for age, I include a quadratic polynomial in age.

The net worth measure is based on a set of eight nominal intervals (see Figure 1 for an example and note that the intervals expand with inflation). I take the midpoint of the interval, deflate using the annual consumer price index, and apply a log transformation.

The sample used to construct each measure is every lawyer-year observation for lawyers currently aged 45-55.[59] The age restriction is designed to prevent newer schools with younger alumni from being unduly penalized.

In addition to these two cardinal measures, I obtained ordinal tiers of law schools from Arewa et al. (2014) in order to provide some external validation. The authors' goal is to establish a classification of school *eliteness* that captures persistent differences in schools with a focus on the middle of the 20th century. They provide seven categories on page 68, and I have added two more categories: one for schools that were too small to be listed in their study, and one for lawyers who reported no school in the MH data.[60] Figure 17 plots log net worth against A ratings, color-coded by the 9 external tiers. The measures are both highly consistent with the external rankings, and seem to complement each other quite well.[61]

Net worth ratings do a very good job of separating the lower half of schools. However, net worths are topcoded and only available for lawyers in smaller cities and towns, so it unsurprisingly does a poor job of separating the top half of schools from each other. Where this measure fails, A ratings succeed. Only about a fifth of lawyers receive an A-rating, so the share of A-ratings essentially captures how many stars a school produces. This is where top schools like Harvard outperform good schools like the University of Minnesota. I produce a final score for law school quality, LSQ, by normalizing each measure into a Z-score and taking a simple average.

[59] As opposed to having one observation per career, this sampling frame allows the *speed* at which lawyers obtain A ratings, which varies considerably, to also influence a school's score.

[60] By the 1930s, firms would seldom consider hiring lawyers who had not attended law school, despite the fact that their own senior partners had often not gone to law school themselves, because it had not been considered essential at the time that they began practicing.

[61] The main exception to this is New York University (NYU), a school with average scores on both measures that Arewa et al. (2014) put in their top tier. They explicitly mention NYU as being a unique case whose placement in the top tier is based more on its recent performance (see footnote 331 on page 68)

Ranking firms. With the LSQ measure in hand, the second step of the procedure forms an index of colleagues' characteristics based on how they predict an individual's own LSQ. The LSQ of a lawyer's colleagues is a very strong predictor of their own LSQ, having a raw correlation of about 0.665, so an obvious starting place is to condition on this variable. My goal is to estimate an equation of the following form.

$$\tilde{\theta}_{i,f,t} = E[LSQ_i|\mathbf{x}_{i,f,t}] = f(\mathbf{x}_{i,f,t})$$

The index $f(\mathbf{x}_{i,f,t})$ is the basis for ranking firms. The simplest possible method would be to assume that $f(\mathbf{x}_{i,f,t})$ is simply an affine function of colleagues' mean LSQ. At the other end of the spectrum, I could incorporate an arbitrary set of characteristics in $\mathbf{x}_{i,f,t}$ and estimate this function non-parametrically. I view this latter method as ideal, but for now I simply choose a relatively small set of characteristics and estimate $f(\cdot)$ as a fully-interacted second-order polynomial. The characteristics $\mathbf{x}_{i,f,t}$ include the number of colleagues, their average law school quality, their average tenure within the firm, their average experience, the share that are A rated, and the population size of the location. Each lawyer's raw index is then transformed into a ranking among all other lawyers working at firms in the same year.

To validate this method, I show that estimated firm ranks are powerful predictors of career success. I consider three outcomes: log rent, log net worth, and whether a lawyer ever obtains an A rating. All three outcomes are strongly predicted by firm rank, conditional on a lawyer's own LSQ, as shown in Table 6.

$$r_{i,f,t} = \frac{1}{N_f} \sum_{j=1}^{N_t} \mathbf{1}\left(\widehat{LSQ}_{j,f,t} < \widehat{LSQ}_{i,f,t}\right) \tag{11}$$

Because individuals in the same firm technically have different colleagues, they will often be measured as having different ranks. Although mildly counterintuitive, this is a small price to pay in order to avoid the mechanical biases that would arise from including an individual's *own* information in the measurement of their firm's rank.

The estimated firm ranks appear to correlate meaningfully with measures of success other than law school. Conditional on your own law school, working at a higher-ranking firm has large positive effects on predicted home-values, net worth estimates, ability ratings, and predicted wages (conditional on being a wage earner). My interpretation of these facts is not that being at a higher ranked firm *causes* you to succeed, but rather that higher-ranking firms select on other correlates of talent besides law school, which are ultimately reflected in these outcomes.

Figure 17: Cardinal measures of law school quality

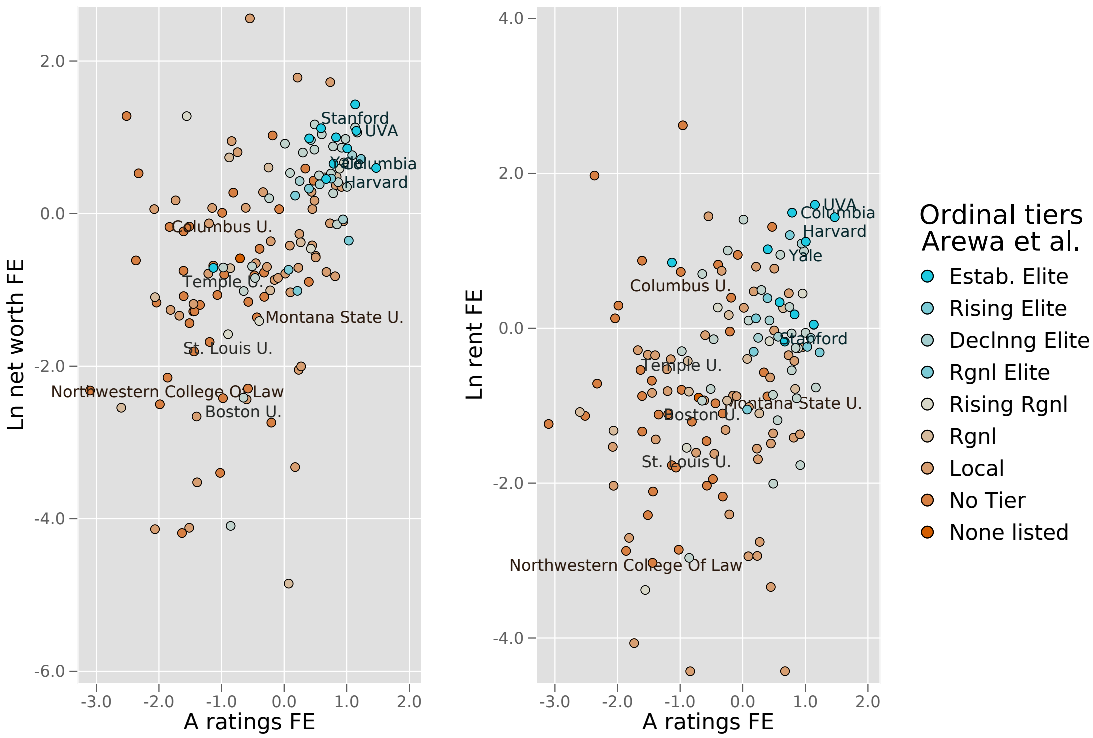


Table 6: Career success vs. firm rank

	Ln 1940 rent	Receives A-rating	Ln net worth
Firm rank	0.347***	0.224***	0.319***
	(0.0154)	(0.00825)	(0.0128)
LSQ	0.107***	0.119***	0.243***
	(0.00648)	(0.00350)	(0.00510)
Mean dep. var.	3.884	.359	12.752
Mkt. size ctrls.	YES	YES	YES
Age ctrls.	YES	YES	YES
Time ctrls.	N/A	YES	YES
N	29,383	45,164	90,417
R^2	0.187	0.083	0.122

Mkt. size, age, and year controls each contain quadratic polynom.
Robust std. errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Acemoglu, Daron and Jorn-Steffen Pischke (1998), "Why do firms train? theory and evidence." *The Quarterly Journal of Economics*, 113, 79–119, URL http://www.jstor.org/stable/2586986.

Akerlof, George A. (1970), "The market for "lemons": Quality uncertainty and the market mechanism." *The Quarterly Journal of Economics*, 84, 488–500, URL http://www.jstor.org/stable/1879431. 10.2307/1879431.

Albouy, David and Mike Zabek (2016), "Housing inequality." Technical report.

Altonji, Joseph G. and Charles R. Pierret (1998), *Employer Learning and the Signalling Value of Education*, 159–195. Palgrave Macmillan UK, London, URL https://doi.org/10.1057/9780230377974_8.

Altonji, Joseph G. and Charles R. Pierret (2001), "Employer Learning and Statistical Discrimination*." *The Quarterly Journal of Economics*, 116, 313–350, URL https://doi.org/10.1162/003355301556329.

Arewa, Olufunmilayo B, Andrew P Morriss, and William D Henerson (2014), "Enduring hierachies in american legal education." *Ind. LJ*, 89, 941.

Baker, George P. and Rachel Parkin (2006), "The changing structure of the legal services industry and the careers of lawyers." *North Carolina Law Review*, 84, 1635.

Banks, Jeffrey S. and Joel Sobel (1987), "Equilibrium selection in signaling games." *Econometrica*, 55, 647.

Bernhardt, Dan (1995), "Strategic promotion and compensation." *The Review of Economic Studies*, 62, 315–339, URL http://www.jstor.org/stable/2297807.

Bidwell, Matthew, Shinjae Won, Roxana Barbulescu, and Ethan Mollick (2015), "I used to work at goldman sachs! how firms benefit from organizational status in the market for human capital." *Strategic Management Journal*, 36, 1164–1173, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.2272.

Braga, Breno (2018), "Earnings dynamics: The role of education throughout a worker's career." *Labour Economics*, 52, 83 – 97, URL http://www.sciencedirect.com/science/article/pii/S0927537118300216.

Burdett, Kenneth and Dale T. Mortensen (1998), "Wage differentials, employer size, and unemployment." *International Economic Review*, 39, 257–273, URL http://www.jstor.org/stable/2527292.

Eeckhout, J. and P. Kircher (2011), "Identifying sorting–in theory." *The Review of Economic Studies*, 78, 872–906.

Farber, Henry S. and Robert Gibbons (1996), "Learning and Wage Dynamics*." *The Quarterly Journal of Economics*, 111, 1007–1047, URL https://doi.org/10.2307/2946706.

Ferreira, Daniel and Radoslawa Nikolowa (2019), "Chasing lemons: Competition for talent under asymmetric information." *European Corporate Governance Institute - Finance Working Paper*. Available at SSRN: https://ssrn.com/abstract=2572036 or http://dx.doi.org/10.2139/ssrn.2572036.

Franck, Thomas W. (2017), "Post-harvard plans." *The Harvard Crimson*, URL https://features.thecrimson.com/2017/senior-survey/after-harvard-narrative/index.html.

Galanter, Marc and Thomas Palay (1993), *Tournament of Lawyers (Paper)*. University of Chicago Press, URL https://www.ebook.de/de/product/3640986/marc_galanter_tournament_of_lawyers_paper.html.

Garicano, Luis and Thomas N. Hubbard (2005), "Hierarchical sorting and learning costs: Theory and evidence from the law." *Journal of Economic Behavior & Organization*, 58, 349 – 369, URL http://www.sciencedirect.com/science/article/pii/S0167268105001551. Theories of the Firm.

Gibbons, Robert and Lawrence F. Katz (1991), "Layoffs and lemons." *Journal of Labor Economics*, 9, 351–380, URL http://www.jstor.org/stable/2535075.

Gourieroux, C., A. Monfort, and E. Renault (1993), "Indirect inference." *Journal of Applied Econometrics*, 8, S85–S118, URL http://www.jstor.org/stable/2285076.

Greenwald, Bruce C. (1986), "Adverse selection in the labour market." *The Review of Economic Studies*, 53, 325.

Haltiwanger, John, Henry Hyatt, and Erika McEntarfer (2018), "Who moves up the job ladder?" *Journal of Labor Economics*, 36, S301–S336.

Kahn, Lisa B (2013), "Asymmetric information between employers." *American Economic Journal: Applied Economics*, 5, 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, 1575–1613, URL https://doi.org/10.1093/restud/rdu021.

Krashinsky, Harry (2002), "Evidence on adverse selection and establishment size in the labor market." *ILR Review*, 56, 84–96, URL https://ideas.repec.org/a/sae/ilrrev/v56y2002i1p84-96.html.

Li, Jin (2013), "Job mobility, wage dispersion, and technological change: An asymmetric information perspective." *European Economic Review*, 60, 105 – 126, URL http://www.sciencedirect.com/science/article/pii/S0014292113000159.

Lloyd K. Garrison, Mitchell Dawson Isidor Lazarus Arthur G. Logan John M. Niehaus Jr. W. A. Roberts, John Kirkland Clark (1938), *The Economics of the Legal Profession*. American Bar Association.

McFadden, Daniel (1989), "A method of simulated moments for estimation of discrete response models without numerical integration." *Econometrica*, 57, 995–1026, URL http://www.jstor.org/stable/1913621.

Moscarini, Giuseppe and Fabien Postel-Vinay (2013), "Stochastic search equilibrium." *The Review of Economic Studies*, 80, 1545–1581, URL http://www.jstor.org/stable/43551567. 10.2307/43551567.

Moscarini, Giuseppe and Fabien Postel-Vinay (2017), "The relative power of employment-to-employment reallocation and unemployment exits in predicting wage growth." *American Economic Review*, 107, 364–368.

Postel–Vinay, Fabien and Jean–Marc Robin (2002), "Equilibrium wage dispersion with worker and employer heterogeneity." *Econometrica*, 70, 2295–2350, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2002.00441.x.

Schonberg, Uta (2007), "Testing for asymmetric employer learning." *Journal of Labor Economics*, 25, 651–691, URL https://EconPapers.repec.org/RePEc:ucp:jlabec:v:25:y:2007:p:651-691.

Sorkin, Isaac (2018), "Ranking firms using revealed preference." *The Quarterly Journal of Economics*, 133, 1331–1393.

Spurr, Stephen J. (1990), "Sex discrimination in the legal profession: A study of promotion." *ILR Review*, 43, 406–417, URL https://doi.org/10.1177/001979399004300405. Doi:10.1177/001979399004300405.

Swaine, Robert T. (1948), *The Cravath Firm and its Predecessors*, volume 2. Ad Press, Ltd., New York.

Waldman, Michael (1984), "Job assignments, signalling, and efficiency." *The RAND Journal of Economics*, 15, 255.

Waldman, Michael (2016), "The dual avenues of labor market signaling." *Labour Economics*, 41, 120 – 134, URL http://www.sciencedirect.com/science/article/pii/S0927537116300185. SOLE/EALE conference issue 2015.

Wilkinson, R. K. (1973), "The income elasticity of demand for housing." *Oxford Economic Papers*, 25, 361–377, URL http://www.jstor.org/stable/2662319.