The Inverted Job Ladder in Skilled Professions

Rustin Partow *

University of California at Los Angeles

JOB MARKET PAPER

This version: February 12, 2020

[Please find the latest version here]

Abstract

How do workers initially match with firms, and how do these matches improve over time? A large job ladder literature devoted to this question proposes a unanimous surplus-ranking of firms in which poached workers move to better firms while displaced workers move to worse firms. Using a new historical dataset on lawyers, I rank law firms based on where their lawyers went to school, finding that poached lawyers move to worse firms while displaced lawyers move to better firms. Guided by these and several other stylized facts, I propose an alternative theory to the standard job ladder approach. In my model, each worker's surplus-maximizing firm assignment is a function of her talent. Incumbent firms privately learn how talented their workers are, and thus only allow adversely selected worker types to be poached. In equilibrium, poached workers therefore move down in rank (to firms where they are more productive). Meanwhile, workers who are retained are revealed over time to have been under-placed, so random displacement shocks move them up in rank (to firms where they are more productive) by temporarily removing the adverse selection problem. The model is well-suited for quantifying the value of labor market institutions that publicly certify talent. By estimating the model, I find that more than 20% of output is lost to misallocation induced by informational frictions. Pre-job market screening devices can substantially raise average earnings, creating a rationale to regulate the timing of job recruitment to prevent it from disrupting informative competition in academics.

*Website: https://rustinpartow.github.io. E-mail address: rustin.partow@gmail.com. I am deeply indebted to my advisors, Maurizio Mazzocco and Moshe Buchinsky. I also thank the following colleagues for useful conversations: Moritz Meyer-ter-Vehn, Dora Costa, Jeremy Lise, V. V. Chari, Sumit Shindu, El Hadi Caoui, Bernardo Silveira, Simon Board, Diego Zuñiga, Alexandre Fon, and Victoria Barone. All errors are my own.

1 Introduction

In markets for skilled professionals, firms can often be ranked by the scale and complexity of their work. For the top firms, who hire engineers to design skyscrapers or hire lawyers to litigate complex bankruptcies, the mistakes of mediocre workers are prohibitively expensive. Meanwhile, firms at the bottom perform easier tasks for which acquiring top talent is unnecessary. Thus, there is an economic return to matching skilled professionals to work that is commensurate with their talent. Some of this return materializes immediately when new graduates are matched into firms based on their academic pedigree. However, as long as academic pedigree is an imperfect signal of true talent, fully efficient assortative matching between workers and firms will require reallocation. How efficient are the initial allocations of workers to firms? And how much does reallocation improve on this?

In the context of the aggregate labor market, a highly successful framework for dynamically assessing labor market efficiency is known as the *job ladder*. However, there are both empirical and theoretical reasons why this framework may be ill-suited to the skilled professions, specifically. This paper's main contribution is to present an alternative framework for studying labor market efficiency, called the *inverted* job ladder, which captures the informational and technological features that make skilled professions unique. I accomplish this in three steps. First, I develop a new historical dataset on lawyers in order to show that job-changing lawyers move down the ladder to worse firms, while lawyers displaced by the dissolution of their firm move up the ladder to better firms. These patterns are the inverse of what standard job ladder models predict. Second, I present a model of dynamic labor market assignment that can explain why the job ladder in law is inverted by using assumptions that are both validated by the data and seem to apply broadly to the skilled professions. Third, I structurally estimate the model in order to assess labor market efficiency 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 is to hire a worker whose location in the talent distribution matches its location in the job ladder. If workers' talents were immediately known, then the model would feature immediate perfectly assortative matching. Instead, each worker enters the labor market with an initial *resumé* of academic feats that signals a distribution of underlying talent and which determines her initial placement. The incumbent employers with whom the worker places privately learn her true talent by 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 successful poaching firm must be lower-ranked than the incumbent.

Each worker's resumé evolves over time, encoding the market's inference from her 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 were allowed to optimally reallocate her without disrupting the future flow of information. This is an indirect way of measuring the quality of the labor market's endogenous learning process, because any remaining inefficiency under this exercise would be due to the social planner's lack of full information. Both components are large, but the informational component is larger. 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 initially *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 increasingly prevalent phenomenon where students are recruited into full-time jobs significantly before they graduate. Early recruitment can disincentivize the effort of recruited students and thus undermine the signaling content of subsequent academic competition. I therefore imagine a policy which reinvigorates academic competition by delaying recruitment, and I use the model to estimate its potential effects on ex ante expected earnings. I find that such a policy is likely to increase average earnings by several percentage points or more, depending on the signaling content of the reinvigorated 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).

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

[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 otherwise might 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.

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 (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

[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.

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 assumptions 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

[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 possible strategic information transmission created by different job titles when using data from this period. Future iterations of my model could certainly incorporate this feature.

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

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.

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

[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).

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 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

[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.

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 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

[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.

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 sample period were identified simply as "members," or "partners."[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 four or more lawyers, is below the age of 55, and entered the market after the year 1931. These are the same sample restrictions that will used in the estimation in Section 5.[15]

[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.

[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

The meanings of the *transitions* variables are described in a few paragraphs 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 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]

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.

[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]In the data, a typical break-up involves a splintering off into different firms, with some colleagues

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 several factual claims that will motivate the theory of the inverted job ladder. The first fact is that lawyers assortatively match into firms based on the quality of their law school. The second fact is that poached workers move to worse firms, and the third fact is that displaced workers move to better firms, which are jointly the opposite of what the usual job ladder framework would predict. The fourth and final fact is that poached workers are adversely selected on plausibly 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.

possibly exiting the market. It is relatively uncommon for entire groups of colleagues to be absorbed by a larger firm. However, in historical biographies of some of the larger firms, there are occasional mentions of absorption of smaller firms in order to expand into new practice areas.

[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.

To establish each fact, I will 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.

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: Assortative Matching by Quality of Law School. I will now show that lawyers assor-

[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.

tatively match into firms based on where they went to school. To do this, I will regress a lawyer's own law school quality measure on the size of her firm and the average law school quality of her colleagues. To avoid a mechanical finding of sorting, average law school quality will be a leave-out mean that omits the individual's own LSQ. The results are presented in Table 2, and reveal that larger firms with a stock of lawyers from better average schools tend to recruit new lawyers who are from better schools. To facilitate interpretation, note that a one-unit increase in LSQ is associated with a 20% increase in predicted housing expenditure.

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. The incentive for firms to specialize in distinct levels of talent may be a unique feature of the skilled professions.

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. Standard job ladder models do not have this crowding-out effect.

Facts 2 and 3: Poaching leads down the ladder, displacement leads up the ladder. I will use average LSQ and firm size to rank firms and study how poached and displaced workers move through the ranks. In order to choose weights for the two measures of firm quality, I will simply use the coefficients estimated in the previous regression. Thus, the principal is that better firms are firms which, based on their characteristics, are predicted to recruit from better schools.[25]

In Table 3, I have regressed changes in firm rank on mutually exclusive indicators for poaching, displacement, and retention (the omitted category). The first column shows that poached lawyers lose an average of 6 percentage points in rank, while displaced lawyers gain an average of 3.7 percentage points. However, poaching and displacement rates are somewhat positively and negatively correlated with firm rank. One concern is that the results are driven by mean-reversion. Hence, the second column controls for the rank of the original firm. The results show that a top-ranked lawyer is predicted to lose 7 percentage points more in rank than a bottom-ranked lawyer—so there is some mean reversion. However, the coefficients on the poached and displaced indicators are virtually unchanged. The third column controls for a host of other potentially important factors, like the quality of the lawyer's law school, her market size, current experience, age, and year fixed effects. These additions increase explanatory power but have minimal effects on the effects of poaching and displacement.

The evidence suggests quite robustly that poached lawyers lose rank, and displaced lawyers gain rank. 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.

Fact 4: Poached workers are adversely selected, displaced workers are not positively selected. Why do poached lawyers move to worse firms? To shed light on this, I will now present evidence that poached lawyers (who lose rank) are adversely selected from those who initially joined a given firm. Meanwhile, displaced lawyers (who gain rank) are not, as one might expect, positively selected. 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

[25] This principal for ranking firms will be consistent with the theory in Section 4.

(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).[26] 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 estimate the probabilities that a lawyer receives an A rating during her career as a function of whether she is poached, displaced, or retained, as well as other relevant aspects of her job history. If an employee who was poached is predicted to have a lower chance of receiving the A rating than an otherwise similar employee who was retained, than I would conclude that the poached employee is likely to have been adversely selected on a latent variable that is correlated with the same talents being judged by the ability rating.

To avoid sample-selection bias, I will take care to estimate these probabilities on a sample of lawyers who are known to continue working for law firms in the data for at least 12 years, and were thus clearly eligible for consideration for an A rating. To ensure that the revelation of each rating outcome was not itself endogenous to the job transitions of interest, I will only examine these individuals during their first six years in the market. Table 4 contains the results of three linear probability estimates, which all suggest that poached lawyers are, ceteris paribus, 4-5 percentage points less likely to receive an A rating than retained lawyers. However, displacement does not appear to carry any such negative association with the attainment of A ratings. If anything, displacement appears to be mildly negative relative to retention. All three specifications assign fixed-effects to each firm-rank quartile, which reveal that lawyers at higher-ranked firms are much more likely to obtain A ratings, even after controlling for the lawyer's own law school quality. The results are also robust to controlling for market size, year fixed effects, and age.

The fact that poached lawyers are adversely selected suggests that firms may 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

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

to hire less talented lawyers. This intuition will be captured by the model.

4 A Model with an Inverted Job Ladder

I now present a model where firms dynamically compete to hire a single worker whose talent is imperfectly known and privately learned by her employer(s). A worker enters the labor market with an academic resumé that noisily signals her underlying talent. The worker initially matches with a firm, the 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.

Poaching and selective retention repeat 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 unfolding history, reflecting some, but not all, of the private information that past employers held when they made their decisions. For exposition's sake, it is best to assume that the worker is just as ill-informed about her talent as the rest of the market. But the results are unchanged by allowing for private information on the worker's side.

Returns to assortatively matching firm and worker types ensure that some endogenous separations occurs in spite of the winner's curse.[27] Poaching offers come from firms who are lower in the ladder than the incumbent, and whose comparative advantage in the incumbent's lemons are so strong that they are rationally interested in paying higher wages to them than the incumbent is willing to pay for its marginally retained worker. This is why the market exhibits downward-directed poaching. Meanwhile, each time that a worker is retained, her resumé improves, creating a building tension where the market knows that she is under-placed but the winner's curse continues to inhibit upward mobility. Displacement shocks resolve this tension by temporarily 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 pri-

[27] In Greenwald (1986), endogenous separations also occurs in spite of the winner's curse, and in spite of firms being homogeneous. In his model, a constant rate of exogenous turnover, which is indistinguishable from endogenous turnover, makes it attractive for below-average workers to quit and *blend in* rather than remain at the incumbent and receive low wages. My model does not have this particular mechanism because exogenous shocks leading to turnover can be correctly distinguished.

vate 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. For exposition's sake, I assume that along the equilibrium path, the worker's information set is identical to the market's information set (her resumé). In fact, this assumption is unnecessary, and even if the worker began with or acquired private information, equilibrium play and payoffs would operate *as if* there were no private information.[28]

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* $[z_{1,1}, z_{1,2}]$ that signals the distribution of her talent. Her actual talent, z, is drawn uniformly from $[z_{1,1}, z_{1,2}]$ and is initially unknown to all players.[29,30] Agents seek to maximize expected earnings or profits. There is no time discounting.

In each period, the worker is attached to an incumbent firm of type θ, with $\theta = \emptyset$ corresponding to a currently unattached worker. The incumbent firm (if one exists) privately knows the worker's talent, z. A period has three stages. In the first stage, the other firms in the market, *poachers*, simultaneously make public spot-wage offers. In the second stage, the incumbent firm privately makes a counteroffer to the worker.[31] [32]

In the third stage, the worker chooses from among the set of available offers. She can be retained at the incumbent, or poached by one of the competing firms. The winning firm pays the worker the promised wage, obtains net output $y(\theta, z)$, and learns her talent. The market updates its priors about the worker based on the outcome of stage three. .[33] With

[28]This result is proved formally, and is probably somewhat dependent on the assumption that incumbent firms make take-it-or-leave-it offers. Under alternative bargaining mechanisms, a worker's private information might improve her bargaining position. In this case,

[29]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.

[30]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.

[31]The assumption that incumbents respond sequentially to poaching offers is a common modeling choice, but there are important exception such as Li (2013) and Greenwald (1986).

[32]By assuming that the counteroffer is private, I avoid an analysis of whether and how an incumbent firm might use the counteroffer to signal the quality of a worker who it did not intend to retain. However, understanding why firms may be induced to write honest recommendation letters about departing employees, especially when firms are long-lived, is an interesting avenue for future research.

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

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—or, equivalently, it forgets its private information about z.[34] If it is the final period $t=T$, then of course the game ends with probability 1.

Net output. The net output function captures revenue net of all opportunity costs other than the wage paid to the worker. Importantly, this introduces the possibility that higher-ranked firms with an *absolute* productivity advantage may nonetheless have a *comparative* disadvantage in hiring low-quality workers who fail to fully leverage the firm's scarce resources and opportunities. Rather than explicitly modeling these opportunities costs, I treat them as a model primitive.

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, 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.[35] 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. $\theta^\star(z)=z$.

An example of a function satisfying these criteria would be the simple linear-quadratic function, $y(\theta,z)=\theta z-\frac{1}{2}\theta^2$. In this case, the first term is a simple Cobb-Douglas revenue function, and the second term is the firm's opportunity cost.

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. $y^{\mathrm{FIM}}(z_{1,t},z_{2,t})=\frac{1}{z_{2,t}-z_{1,t}}\int_{z_{1,t}}^{z_{2,t}}y(z,z)dz$ is the worker's *expected full information output*.

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

[35] 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.

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: (1) poachers make wage offers, (2) the incumbent makes a counteroffer, and (3) the worker chooses an offer. In order to build intuition, I will start with a one-period model and 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) offer acceptance rules such that the beliefs are consistent with Bayes' rule, whenever it applies, and the wage offer and offer acceptance rules are sequentially rational. I will now introduce a refinement on off-equilibrium-path beliefs.

Assumption 3 (Off-path beliefs)**.** 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 $z = z_2$.

Assumption 3 is in the spirit of the Divinity Criterion of Banks and Sobel (1987). It 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 outcome *most* profitable from the standpoint of the incumbent firm and the worker. This restriction on off-path beliefs will render all essential features of equilibrium behavior unique. A similar result could be achieved through the introduction of infrequent idiosyncratic shocks to the incumbent or worker's affinity for continuing the match.

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. Once employed at an incumbent firm, the worker is partially insulated from outside wage competition due to the incumbent's private information. The incumbent uses a cutoff rule when deciding whether or not to retain the worker, in which case she is given a counteroffer that makes her indifferent to being poached. Poachings will truncate the resumé down, retentions will truncate the resumé up, and displacements will have no effect on the resumé. Being displaced is always payoff enhancing from the worker's perspective, because it frees her from the winner's curse and allow her to receive more competitive poaching offers.[36]

The one-period model. Let us quickly explore what happens in a one-period version of the model, when $t = T$. In this case, the incumbent firm is treated as exogenous. 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 will match the best offer, if and only if paying this wage to the worker and obtaining output from her would be profitable. As is often the case, in equilibrium the worker must behave as if she prefers the incumbent in cases of ties, because otherwise the incumbent could always pay an infinitesimally small premium to break the tie.

Let θ_T denote 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

[36]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.

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 58.

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.

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 59.

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 makes a counteroffer. In stage 3, the worker chooses to either remain at the incumbent, or to separate and accept one of the poaching offers.

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 2.

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 resumé and 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 a worker who has resumé $[z_1, z_2]$ and matches with a firm of type θ at the end of stage 3.

Definition 3. The *value function* is $V_t(z_1, z_2) = \max_\theta V_t^P(\theta, z_1, z_2)$. The *firm assignment function*, $\theta_t(z_1, z_2)$, is the associated maximizer.

Definition 4. The *cutoff rule function*, $\zeta_t(\theta, z_1, z_2)$, 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.

Upon observing the incumbent firm's counteroffer, the worker will privately surmise whether or not she is above or below the cutoff. In theory, if she were to separate from the incumbent firm, her beliefs would diverge from those of the poaching firms, who would believe the worker is below the cutoff. We might be concerned that the worker's own beliefs could, in these cases, become relevant to actions or payoffs and thus require us to keep track of additional variables. The next lemma shows how the assumptions of the model explicitly rule this out. Assumption 3 plays a key role in this by ensuring that no matter how many times the market observes off-equilibrium path behavior, it will never contradict its previous beliefs. This constrains the potential for the worker to manipulate the market's beliefs, by, for example, taking off-equilibrium path offers and being retained off the equilibrium path, and makes it unprofitable to attempt to do so. Lemma 3 will also imply that the main analysis would be unchanged had we assumed that the worker begins the game with private information about her true talent.

Lemma 3. Suppose the worker has private information about z. Conditional on the resumé and incumbent firm, the worker's payoffs do not depend on her private information.

See proof on page 59.

Definition 5. The *ex post incumbent profit function*, $\pi_t^I(\theta, z_1, z_2, z)$, is the expected net present profits of an incumbent firm of type θ at the beginning of period t when attached to a worker with resumé $[z_1, z_2]$, given that the worker's true talent equals z.

Definition 6. 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 at the beginning of period t when attached to a worker with resumé $[z_1, z_2]$.

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. Realized profits are likely to be positive in states of the world where the worker's talent is above average and negative in states of the world where she is below average. Thus, the ex ante profit function is an ex ante expectation formed based on the information contained in the resumé.

Because firms effectively Bertrand compete for the right to hire the worker, the profits of doing so will be exactly equal to the entry wage paid. Taking this to its logical conclusion, we arrive at the following intuitive result.

Lemma 4. $V(z_1, z_2)$ is equal to the worker's expected output in equilibrium.

Corollary 1. $V(\theta, z_1, z_2)$ is equal to the worker's expected output, given that all future play is in equilibrium.

Another way to understand this result is as follows. Suppose that ex ante a worker is expected to earn less than her output. Then ex ante, some firm must be making positive expected profit, in which case its competitors are failing to outbid it to recruit the worker in states of the world where doing so would be strictly profitable.

As in the one period model, if the incumbent wants to retain the worker, it will counter the poaching wage offers with a wage that makes the worker indifferent between being retained and being poached, and will always win ties. Moreover, thanks to Lemma 3, we know that the communication of the incumbent's information at stage 2, via the counteroffer, does not change the worker's prospects of being poached. Unlike the one-period model, the incumbent's counteroffer will not necessarily be as high as those of the poaching firms. This is because the worker's decision to leave the incumbent would result in a negative inference about her type, referred to as a *marking phenomenon* in Greenwald (1986). The worker prefers the offer that promises the highest expected sequence of wages. Poached workers experience persistently low future wages and must therefore be compensated for this with higher up-front payment.[37]

It will be useful to imagine the different retention wages that would ensure indifference between retention and poaching for different possible cutoffs (only one of which will actually be an equilibrium cutoff).

Definition 7. 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 $[z_1, z_2]$ and attached to incumbent θ indifferent to being poached, and revealed below ζ, and being retained and revealed above ζ.

We can easily characterize the hypothetical retention wage by noticing that if the worker is poached, then her continuation value is the value function evaluated at the resulting downward-truncated resumé. The proof is intuitive and is therefore printed below.

Lemma 5.

$$w_t^R(\theta, z_1, z_2, \zeta) = V_t(z_1, \zeta) - \delta\left((1-\lambda^D)V_{t+1}(\zeta, \zeta_{t+1}(\theta, \zeta, z_2)) + \lambda^D V_{t+1}(\zeta, z_2)\right). \qquad (2)$$

Proof. If the worker is retained, then her resumé will become $[\zeta, z_2]$. With probability λ^D she remains attached to firm θ, and is once again made indifferent to being retained and being poached. The firm's cutoff rule in this event will be $\zeta_{t+1}(\theta, \zeta, z_2)$. Thus, without loss of generality, we may express the worker's continuation value in the event that she is retained and not displaced as $V_{t+1}(\zeta, \zeta_{t+1}(\theta, \zeta, z_2))$.

Meanwhile, if the worker is displaced, then her resumé will remain at $[\zeta, z_2]$ and she will earn continuation value $V_{t+1}(\zeta, z_2)$. Thus, to ensure the worker's indifference to being retained and being poached in the current period, we require

$$w_t^R(\theta, z_1, z_2, \zeta) + \delta\left((1-\lambda^D)V_{t+1}(\zeta, \zeta_{t+1}(\theta, \zeta, z_2)) + \lambda^D V_{t+1}(\zeta, z_2)\right) = V_t(z_1, \zeta),$$

which gives the result, save for a minor algebraic manipulation. □

The cutoff rule is the level of worker talent at which the firm's benefit of retaining the worker—net output plus the future ex post incumbent profit—equals the hypothetical retention wage. If multiple cutoffs satisfy this condition, only the largest can be part of an equilibrium, due to the first-mover advantage of poaching firms which allows them to force coordination on higher cutoffs. This is captured formally below.

[37] Readers who find this counterintuitive should remember that in this model, poached workers are always indifferent to being retained, despite the fact that in equilibrium they are of lower quality.

Proposition 1. The *equilibrium cutoff rule* is

$$\zeta_t(\theta, z_1, z_2) = \sup_{\zeta \in [z_1, z_2]} \left\{ y(\theta, \zeta) + (1 - \lambda^D)\delta\pi^I_{t+1}(\theta, \zeta, z_2, \zeta) - w^R_t(\theta, z_1, z_2, \zeta) > 0, \forall z > \zeta \right\}. \tag{3}$$

See proof on page 60.

Definition 8. The *equilibrium retention wage*, $w^R_t(\theta, z_1, z_2)$, is the wage that a type θ firm pays to retain a worker with resumé $[z_1, z_2]$ along the equilibrium path.

Corollary 2. The equilibrium retention wage is simply the hypothetical retention wage, evaluated at the equilibrium cutoff rule.

$$w^R_t(\theta, z_1, z_2) = w^R_t(\theta, z_1, z_2, \zeta_t(\theta, z_1, z_2)).$$

The profit from retaining the worker is therefore expected net 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.

Lemma 6. For all $t > T$, $\Pi^I_t(\theta, z_1, z_2) = 0$. For all $t \leq T$,

$$\begin{aligned} \Pi^I_t(\theta, z_1, z_2) = & \frac{z_2 - \zeta}{z_2 - z_1} \left(\bar{y}(\theta, \zeta, z_2) + \delta(1 - \lambda^D)\Pi^I_{t+1}(\theta, \zeta, z_2) - w^R \right), \\ & \text{subject to } \zeta = \zeta_t(\theta, z_1, z_2), \\ & \text{and } w^R = w^R_t(\theta, z_1, z_2). \end{aligned} \tag{4}$$

The ex post incumbent profit function can also be described recursively.

Lemma 7. For all $t > T$, $\pi^I_t(\theta, z_1, z_2, z) = 0$. For all $t \leq T$,

$$\begin{aligned} \pi^I_t(\theta, z_1, z_2, z) = & \mathbf{1}(z > \zeta) \left(\bar{y}(\theta, z) + \delta(1 - \lambda^D)\pi^I_{t+1}(\theta, \zeta, z_2, z) - w^R \right), \\ & \text{subject to } \zeta = \zeta_t(\theta, z_1, z_2), \\ & \text{and } w^R = w^R_t(\theta, z_1, z_2). \end{aligned} \tag{5}$$

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 neither exits nor is displaced, and (3) the continuation value if the worker is displaced.

Lemma 8.

$$
\begin{aligned}
V_t^P(\theta, z_1, z_2) = & \underbrace{\bar{y}(\theta, z_1, z_2) + \delta(1-\lambda^D)\Pi_{t+1}^I(\theta, z_1, z_2)}_{\text{initial wage}} \\
& + \underbrace{\delta(1-\lambda^D)V_{t+1}(z_1, \zeta_{t+1}(\theta, z_1, z_2))}_{\text{continuation value of non-displaced worker}} \\
& + \underbrace{\delta\lambda^D V_{t+1}(z_1, z_2).}_{\text{continuation value of displaced worker}}
\end{aligned}
\tag{6}
$$

The first component, the current wage, is competed up to the point of zero ex ante profit, equaling net output plus the future ex ante profit function. The second component captures the fact that, if the worker is not displaced, then regardless of whether she is poached or retained, she receives her outside option of being poached.

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 3, I have illustrated the infinite horizon Markov equilibrium.

The inverted job ladder. I will now 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 2. 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 61.

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 that the trivial case of no poaching is *not* the only case where Proposition 2 applies. In particular, as δ goes to 0, poaching must become positive for all cases where $z_1 < z_2, z_1 < \theta$.

Corollary 3. There exists a range of values for δ for which there is strictly positive poaching 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.

See proof on page 61.

Thus, as long as the discount factor is sufficiently high, the worker's resumé cannot improve so much as to eventually reveal that she is with probability one better than her incumbent firm. This condition is key for predicting that poaching is downward-directed. The basic intuition is that, as workers become sufficiently patient, the stigma or *marking* 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 required remaining at a firm paying extremely 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 is precisely the leverage held by the incumbent firm, and implies that as long as δ is sufficiently large, any particular employment spell beginning at (θ, z_1, z_2) has a cap on the degree to which z_1 can increment up through market updating.

Assumption 4. Assume that δ is 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. They both require the assumption that the equilibrium firm assignment function, $\theta_t(z_1, z_2)$, be increasing in its arguments. It is easy to provide a set of sufficient conditions for this to hold. First, if either the probability of exit or of displacement are sufficiently high, then contemporaneous expected output can be made arbitrarily more important than future output in determining which firm types make the worker the most attractive offer. The supermodularity of contemporaneous output will result in firm assignment increasing in z_1 and z_2 (indeed, assignment is increasing in the one-period model). Second, regardless of exit and displacement probabilities, there is (from Proposition 2) some point where, when $\frac{z_1}{z_2}$ is sufficiently close to 1, equilibrium poaching rates go to 0. Therefore, for workers whose resumés are sufficiently precise, optimal assignment becomes arbitrarily close to static output maximization. Hence, this condition is guaranteed to hold as either $1 - \delta$, λ^D, or $\frac{z_1}{z_2}$ approach 1.

Assumption 5. The equilibrium firm assignment function, $\theta(z_1, z_2)$, is increasing in both of its arguments.

Theorem 1. Consider the T period finite horizon model. Suppose that the equilibrium firm assignment function, $\theta_t(z_1, z_2)$, is increasing in its arguments. Then along every equilibrium path with poaching in time t, the poaching firm is of lower type than the incumbent firm. I.e., $\theta_t(z_1, \zeta_t(\theta, z_1, z_2)) < \theta$ for all (θ, z_1, z_2) along the equilibrium path.

Proof. By Assumption 4, $\zeta_t(\theta, z_1, z_2) < \theta$. In this case, the optimal firm $\theta_t(z_1, \zeta)$ is below $\theta_t(\zeta, \zeta)$, and since $\zeta < \theta$, this is below $\theta_t(\theta, \theta)$, which is equal to θ (a worker with no uncertainty in their type would always match with the firm at which she is most productive). □

Theorem 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.[38] I will identify λ^D by the degree to

[38] 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.

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.

Assumption 6. 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 7. 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 8. Let $\alpha \in (0, \phi)$, $\phi > 0$.

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

The first part of Assumption 6 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 7, 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 8 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.[39]

Assumption 8 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 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.[40]

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

Lemma 9.

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

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 will 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 3 (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

[39] 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.

[40] For interested readers, these are all derived in Appendix A.

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

See proof on page 62.

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. In this case, a separating worker is with probability one less talented than a retained worker with the same prior history—equivalently, A rated lawyers will *never* have experienced separations in their past. On the other hand, if all separation were exogenous, than A rated lawyers would have similar past separation rates to the overall population of lawyers. The reduced probability of separation for A rated lawyers maps directly into the fraction of turnover that is endogenous.

To identify ϕ and γ, I first exploit a homogeneity result linking proportional variation in initial resumés to permanent income.

Lemma 10. For any T-period version of the game, the value function, ex ante incumbent profit function, and equilibrium retention wage are all homogeneous of degree ϕ. The firm assignment function and equilibrium cutoff rule are homogeneous of degree 1.

See proof on page 62.

From the homogeneity result, it follows that ϕ, the returns to scale in net output, is also the elasticity of permanent income with respect to talent, when the uncertainty in talent, measured by $\frac{z_2}{z_1}$, is held fixed.

Corollary 4.

$$\ln V(z_1, z_2) = \phi \ln z_1 + \ln V(1, \frac{z_2}{z_1}).$$

Proof. By homogeneity, we can write $V(z_1, z_2) = z_1^{\phi} V(\frac{z_1}{z_2}, 1)$. Taking logs yields the result. □

The following results will now derive a closed form relationship between permanent income and the probability of obtaining an A rating, conditional on $\mathbf{x}$. Both conditional moments will shown to be strictly increasing functions of $g(\mathbf{x})$, and thus will have a one-to-one relationship with eachother.

Lemma 11. Let $\mathbf{x}$ denote the instrument (such as LSQ) described in Assumption 6. 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 $\mathbf{x}$. Then, for $\mathbf{x}$ in said interval,

$$\tilde{v}(\mathbf{x}) = \phi g(\mathbf{x}) + c_1,$$

$$p_A(\mathbf{x}) = \frac{\gamma}{\gamma - 1}\left(1 - c_2 e^{-g(\mathbf{x})}\right),$$

and

$$\tilde{v}(\mathbf{x}) = c_3 - \phi \ln\left(1 - \frac{\gamma - 1}{\gamma} p_A(\mathbf{x})\right). \tag{7}$$

for some constants c_1, c_2, and c_3.

See proof on page 64.

The nonlinear mapping between permanent income and A ratings in Equation 18 jointly identifies ϕ and γ.

Proposition 4 (Identification of ϕ and γ)**.** 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

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

See proof on page 65.

Thus, ϕ and γ are identified from two distinct measures of curvature in the relationship between expected log earnings, $\tilde{v}$, 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. Numerically, I find that the first-year turnover rate is an increasing function of α when all of the other parameters are held fixed to those estimated previously. However, I do not yet have a proof that this monotonic relationship holds for all possible values of the other model parameters.

Conjecture 1. The first-year turnover rate is monotonically increasing in α, holding the other parameters fixed.

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 6 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 $1 - 6.87\% = 0.931$.

Estimation of λ^D. To estimate λ^D, I use the previous result stating that $\lambda^D = Pr(\text{Separate}|\text{A})$. This allows me to simply estimate the average turnover rate among 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 success for the model, I am unable to reject the hypothesis that turnover rates are identical by year of experience for A rated 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 4. 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 4 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.[41] 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.[42] My 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.[43] 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

$$\tilde{h}_i = \tilde{v}_0 - \phi \ln\left(1 - \tilde{\gamma} p_{A,i}\right) + u_i. \tag{8}$$

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 8 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 8 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

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

[42]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.

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

$$p_{A,i} = \frac{1}{\tilde{\gamma}} \left(1 - ke^{-\frac{\tilde{h}_i - \bar{h}}{\phi}} \right). \tag{9}$$

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 9 with respect to $\tilde{h}_i$ around $\bar{h}$, to get

$$p_{A,i} \approx \underbrace{\frac{1-k}{\tilde{\gamma}}}_{b_0} + \underbrace{\frac{k}{\tilde{\gamma}}\phi}_{b_1} \left(\tilde{h}_i - \bar{h}\right) + \underbrace{\frac{k}{2\tilde{\gamma}}\phi^2}_{b_2} \left(\tilde{h}_i - \bar{h}\right)^2. \tag{10}$$

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 10 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 4, 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

$$p_{A,i,m} = b_{0,m} + b_{1,m} \left(\tilde{h}_i - \bar{h}_m\right) + b_{2,m} \left(\tilde{h}_i - \bar{h}_m\right)^2. \tag{11}$$

I first estimate the equation pooling across counties. Next, I cut the sample into three market-size categories.[44]

The estimates of ϕ and γ are presented in Table 5. 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

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

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 $\gamma = 6.03$.[45]

Having estimated γ, ϕ, and λ^D, I will now estimate the final parameter α using the 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 $[1, \gamma]$,[46] 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

[45] 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.

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

matrix, $W^\star$, is the inverse of the variance-covariance matrix of the moments (evaluated at the true parameter).[47]

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, 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.[48]

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.[49]

5.3 Estimation Results

All parameter estimates (or calibrated values) and standard errors are listed in Table 6. 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

[47]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.

[48]One fortunate aspect of solving for the Markov equilibrium is that one typically does not need to solve for the ex post incumbent profit function. This function, which appears in Proposition 1, is only relevant when evaluated at the marginally retained worker type. As long as the future turnover rate is non-zero, the marginally retained worker is guaranteed to separate next period. Thus, only in those cases where the turnover rate is zero does the ex post incumbent profit function come into play.

[49]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.

largest and smallest estimate by 1.96 to get a "rule of thumb" adjusted standard error.[50]

Results from the simulated model. The simulated and empirical experience turnover profiles are plotted in Figure 4. 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 experience-turnover profile without added refinements. The refinements that would help match a less convex 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. However,

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 5, 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. First, how much less efficient is the market compared to a full information benchmark. Second, how much inefficiency can be attributed to the speed at which the market identifies talent, relative to mismatched assignments that are publicly known to be inefficient? And third, what kinds of policy interventions could 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

[50] 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).

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 6 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.[51] 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 attributable to a lack of information. In particular, very little is learned about mature workers, for whom the mismatch with their current employer is not large enough to induce significant poaching.

6.1 Reinvigorating academic competition through delays in recruitment

Through the lens of the model, technologies that lead to more precise pre-job market public screening of talent will raise expected earnings, all else held equal. Thus, one might

[51] By the first welfare theorem, a competitive market would reach the same static optimization outcomes. The first welfare theorem does not apply to the baseline model due to incomplete information.

naturally ask whether there are welfare-enhancing policies that improve pre-job market screening technologies. One of the more important sources of pre-job market screening, especially for lawyers, is academic competition. After enrolling in law school, law students distinguish themselves by competing to earn good grades. Although this type of competition produces winners and losers, participation is valuable ex ante because it facilitates better subsequent assignment of winners and losers alike.

The bizarre feature of modern academic competition within law schools is the extent to which it is concentrated in the first year. Second and third-year grades are not considered to be nearly as informative as first-year grades. This phenomenon is commonly attributed to a very early recruitment process that prioritizes first year grades. Specifically, the top firms use first-year grades only when deciding to whom they will offer summer internships, which are, roughly speaking, a necessary and sufficient condition for getting a full-time offer. Once the students who perform best in the first year receive their internship offers, their incentives to continue competing in the second and third year decline, reducing the vigor of competition in years two and three.

The second and third years of law school are potentially valuable opportunities to provide pre-job market screening. So why does the market forgo these opportunities? It could be the result of a coordination failure on the part of firms. Socially, the market is better off if recruitment is delayed in order to encourage more academic competition, but privately each firm wants to recruit earlier than its rivals. Thus, it could be in society's interests to regulate the timing of recruitment–i.e., ban recruitment that is earlier than a specific date.[52]

Let us imagine a simple policy which says that law students cannot apply for internships until the spring, rather than fall, of their second year of law school.[53] Assume that by giving employers one additional semester of grades to make their decisions, this policy *reinvigorates* competition in the fall of the second year. A student enters the semester with resumé $[z_1, z_2]$. Assume that her participation in this new competition produces a public pre-job market screening of talent into two categories: those below some cutoff, $\tilde{\zeta}$, and those above it. The student then enters the baseline model as a worker with the resumé produced by the screen.

[52]A skeptic might ask why universities haven't internalized this problem themselves and voluntarily banned early recruitment. I would again suggest a coordination failure. If a single law school banned early recruitment, then the top law firms may prefer to stop recruiting there, rather than redesign their recruitment strategies. At the same time, such a school would have a difficult time explaining the economic virtues of such a ban to prospective students. Hence, unilaterally banning early recruitment could be very costly in the short run.

[53]To the extent that this later recruitment period creates administrative burdens on law firms, I will abstract from them here.

The winners of this new competition earn expected income $V(\tilde{\zeta}, z_2)$ with probability $p^{\text{win}} = \frac{z_2 - \tilde{\zeta}}{z_2 - z_1}$, while the losers earn $V(z_1, \tilde{\zeta})$ with probability $1 - p^{\text{win}} = \frac{\tilde{\zeta} - z_1}{z_2 - z_1}$. The percentage increase in a worker's expected earnings is

$$\Delta^{Earnings} = \frac{p^{\text{win}} V(\tilde{\zeta}, z_2) + (1 - p^{\text{win}}) V(z_1, \tilde{\zeta})}{V(z_1, z_2)} - 1.$$

Notice that $\tilde{\zeta} = z_2 - p^{\text{win}}(z_2 - z_1)$. Because the value function is homogeneous in (z_1, z_2), we can rewrite the above as

$$\begin{aligned}
\Delta^{Earnings} &= \frac{p^{\text{win}} V(z_2 - p^{\text{win}}(z_2 - z_1), z_2) + (1 - p^{\text{win}}) V(z_1, z_2 - p^{\text{win}}(z_2 - z_1))}{V(z_1, z_2)} - 1 \\
&= \frac{p^{\text{win}} V(\gamma - p^{\text{win}}(\gamma - 1), \gamma) + (1 - p^{\text{win}}) V(1, \gamma - p^{\text{win}}(\gamma - 1))}{V(1, \gamma)} - 1 \qquad (12) \\
&= \Delta^{Earnings}(p^{\text{win}}).
\end{aligned}$$

Because we already have knowledge of the equilibrium value function and the parameter γ, the percentage change in earnings under the counterfactual is simply a function of the chance of winning or losing the competition, p^{win}. There are several reasonable values to consider for p^{win}. First of all, what is the winning probability that could most feasibly be offered by an academic competition? For this, I would argue that the natural place to look is the labor market, which through the lens of the model generates binary signals via poaching and retention. The first-year separation rate is about 16%, and the rate of exogenous displacement is estimated to be 7.5%, implying that the first-year poaching rate is 9.5%. Thus, if academic competition were to simply mimic the way that workers effectively compete to be retained, then the winning probability should be 90.5%. In this case, I estimate an increase in expected earnings of about 2%.

The second intuitive winning probability is 50% (under my modeling assumptions, this probability minimizes the expected variance in the distribution of z after the competition, and is in that sense the most informative winning probability).[54] In this case, I find welfare gains of 4%. Finally, we ought to be interested in the winning probability that maximizes expected earnings. One might expect that this probability would be close to the 50% number above. It turns out, however, that the optimal winning probability is 16%, which generates welfare gains of 6%. Why is this the case? As some basic intuition, think about the complementarity between academic screening and the subse-

[54] For reference, bar exams typically pass three-quarters and certified public auditor accountant (CPA) exams pass about one-half of first-time takers.

quent screening that occurs in the labor market. In the labor market, poaching rates are relatively low, and converge to zero over time. Thus, the labor market is very quick to reveal lemons in the left tail of an initial distribution of talent, but is very slow to identify the stars. Hence, it is quite intuitive that an optimal pre-job market screening technology would help identify stars and thus complement, rather than substitute, the endogenous screening of the labor market.

These results provide some basic economic evidence to warrant thinking more deeply about the possible adverse consequences of the early recruitment of skilled professionals—a phenomenon which is not only occurring in law, but also in accounting, finance, and consulting.

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, reduces turnover, and increases the degree of misallocation. Accounting for these forces allows my model to more realistically quantify market efficiency compared to most of the past models of asymmetric information in the labor market.[55]

In skilled professions, a surplus or preference-based ranking of firms is unlikely to be the same for different workers—i.e., more talented workers will rank higher the firms that do more challenging work. Nonetheless, firm rankings feature very saliently in skilled

[55] These forces have been accounted for by recent work on dynamic markets with adverse selection, most recently in Fuchs and Skrzypacz (2019), but I am not aware of any such papers pertaining to the labor market.

professions, and seem quite connected to the perceived success of their employees. So what is the economic content of such rankings? This paper lays the foundation for a job ladder where (1) the top firms are the ones that endogenously recruit the employees with the best observable talents, and (2) poaching is directed down the job ladder. The inverted job ladder proposed in this paper can easily replicate some of the relationships that empirical researchers have come to expect: firms that are higher in the ladder should pay higher average wages and be more productive. 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.

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. 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.

In counterfactual analysis of the estimated model, I underscored how plausible coordination failures could result in premature recruitment that disrupt academic competition, robbing the market of opportunities to screen talent and costing 2%-6% of earnings. That analysis also found that optimally calibrated academic competitions will prioritize the identification of stars (and have very low success rates) in order to complement the tendency of endogenous poaching to identify workers in the left tail of the talent distribution.

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 nursing? Promising avenues for this research include examinations of earnings dynamics and patterns in firm-to-firm transitions, which mirror some of the standard empirical research supporting the standard job ladder, but which are limited to workers in skilled professions.

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, which I anticipate will be key to improve the model's ability to fit empirical turnover profiles. A fourth refinement would be to allow for general human capital accumulation.

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: Graphical rendition of two period game

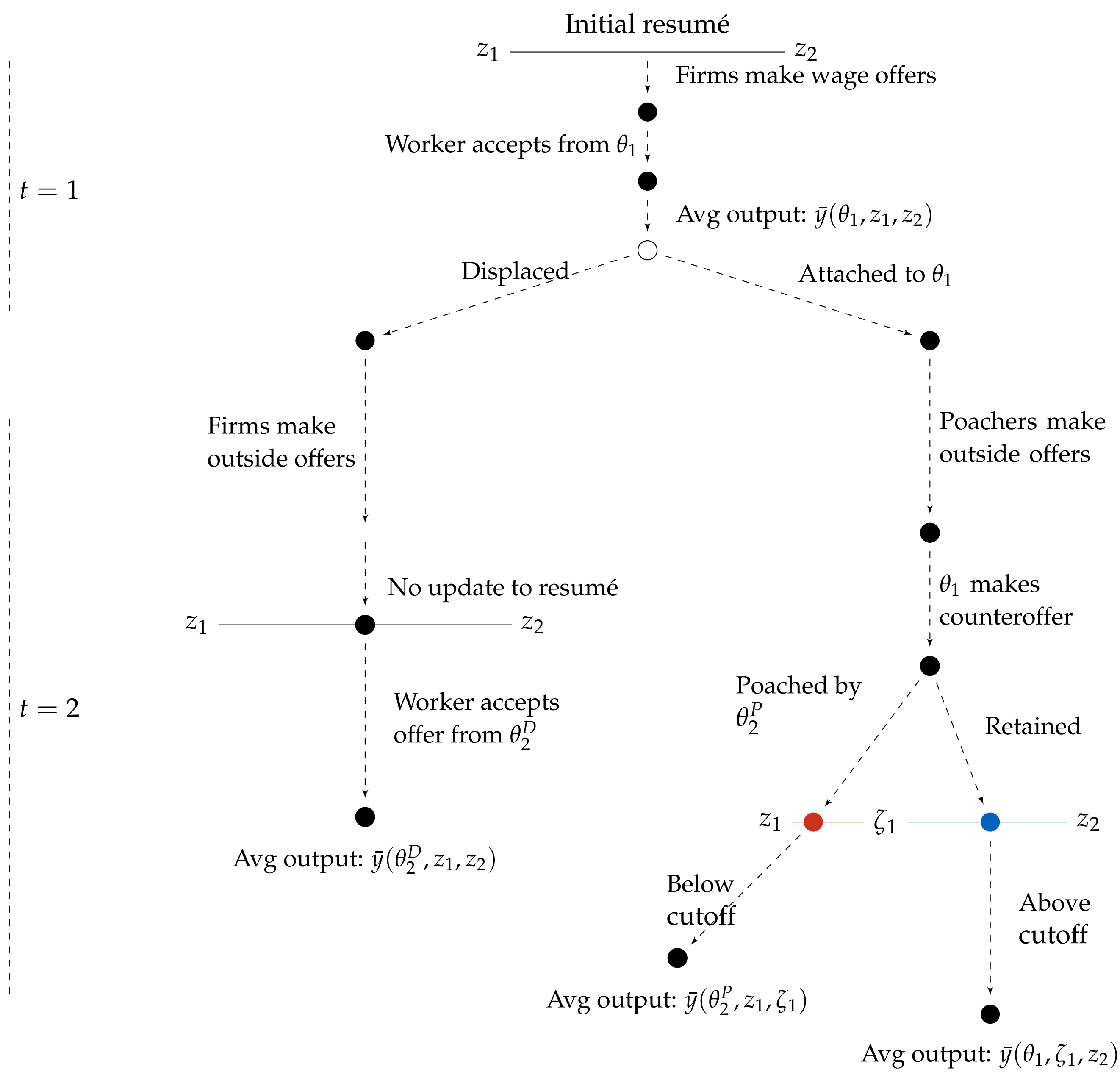

Figure 3: State to state transitions of Markov game

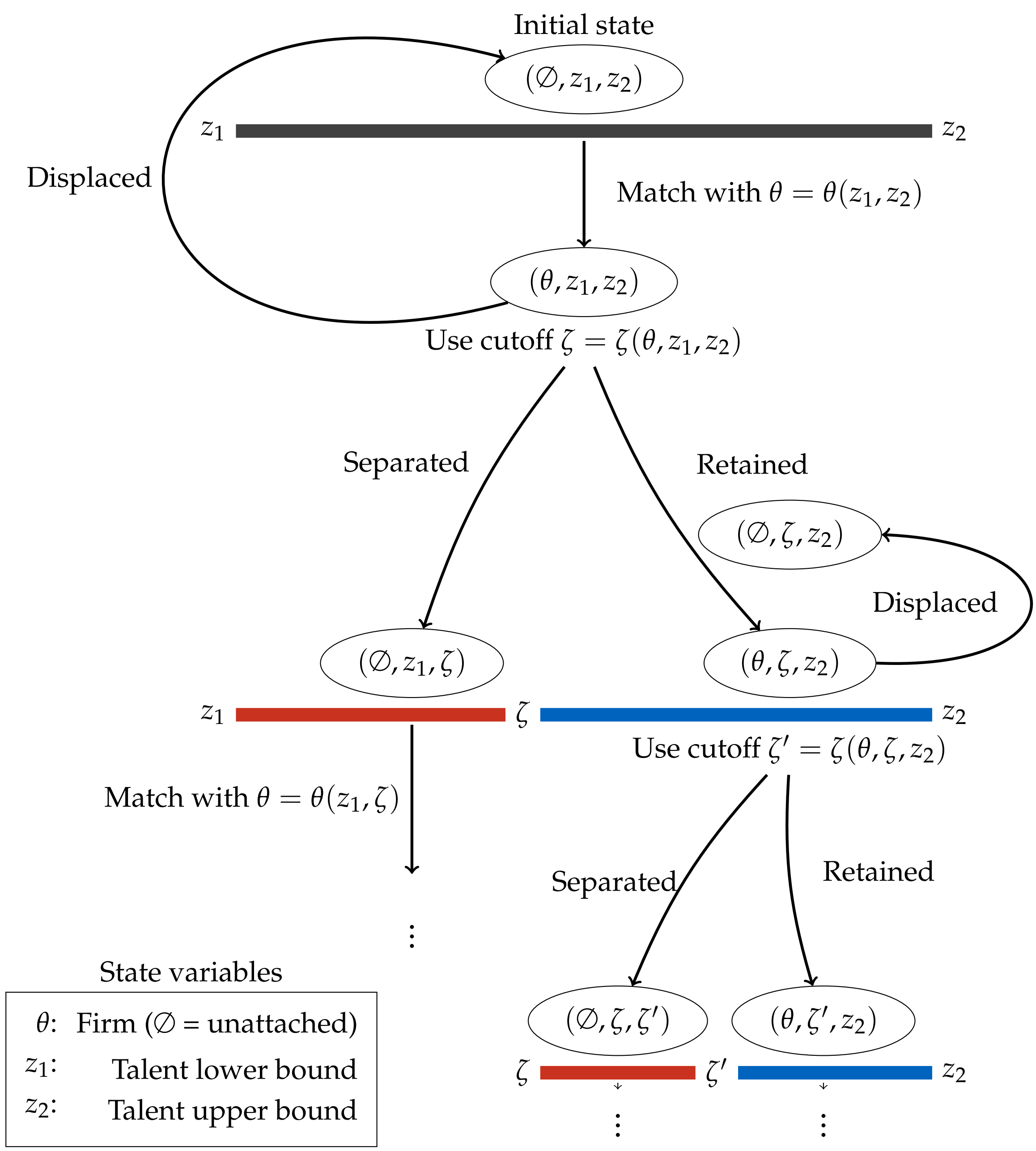

Figure 4: Simulated versus actual turnover

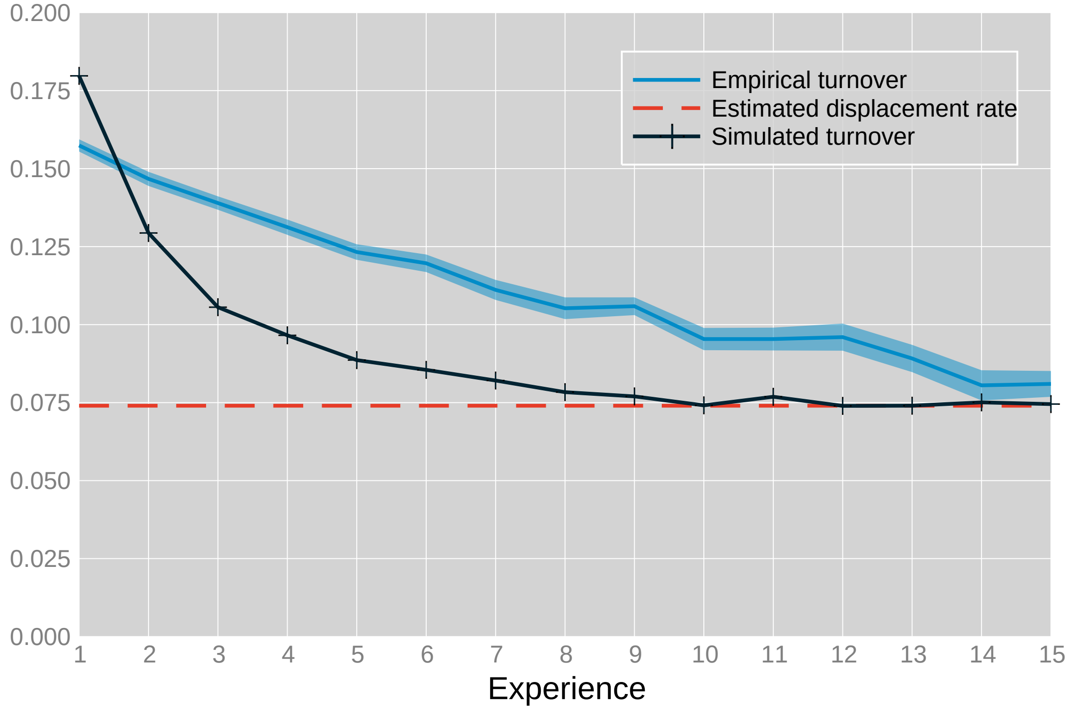

Figure 5: Distortions in equilibrium placement

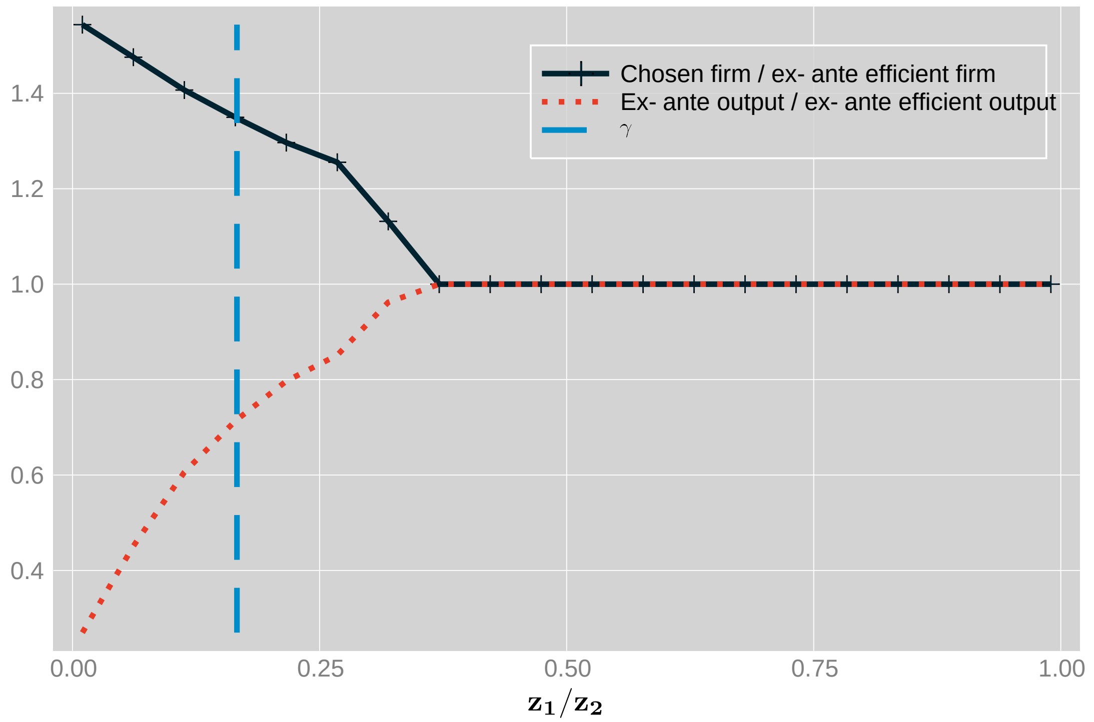

Figure 6: Expected output (% of full-information) of currently unattached worker

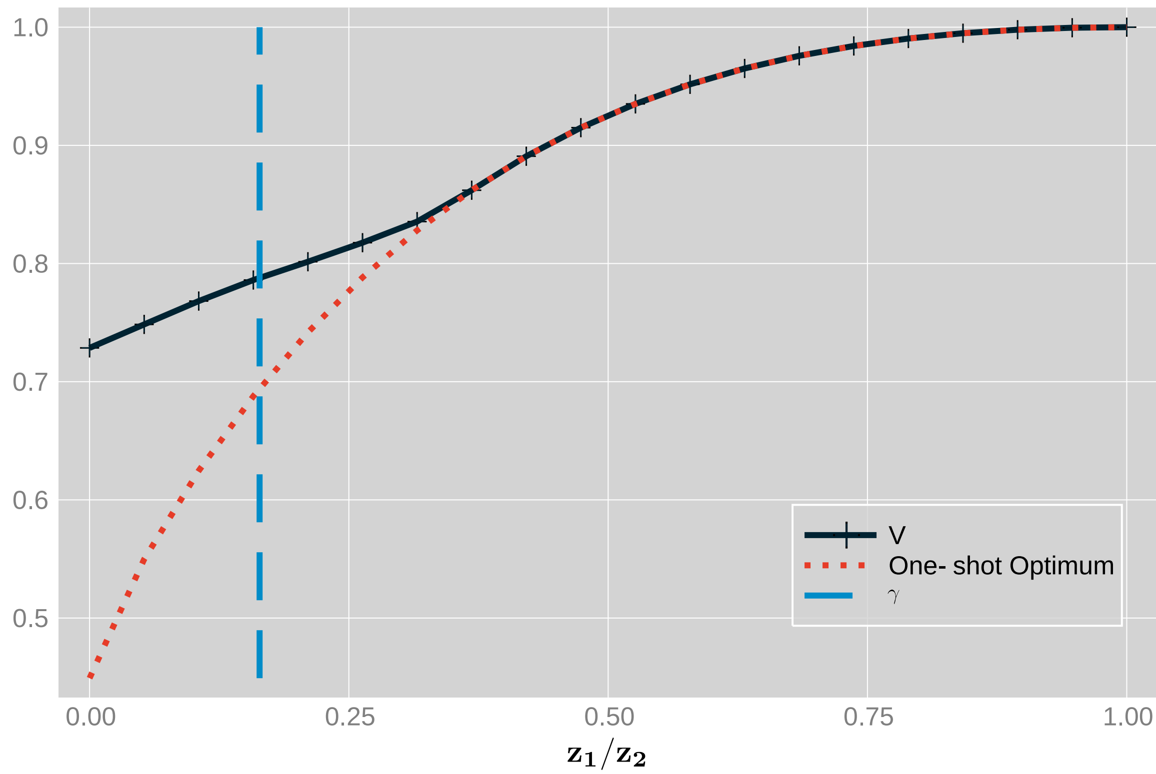

Table 1: Summary statistics by lawyer-year

	Mean	Std.dev.	*p*.05	*p*.95
Age	39.13	7.41	29	52
Exper.	8.71	6.35	1	21
A Rated	0.36			
Mkt. size	5,387.99	7,695.94	28	24,162
Firm size	12.33	13.71	4	42
Poached	0.06			
Displaced	0.03			
Retained	0.83			
Exit	0.08			
Obs.	347,379			

Sample: Lawyers currently in firms of size 4+, 1933-1960, aged 22-55, non-attriting
Mkt. size reflects number of lawyers working in local town or city
ARated computed only on eligible lawyers (10+ years experience)

Table 2: Assortative Matching by Law School Quality

Dependant variable	**LSQ**
Avg LSQ (leave-out)	0.662***
	(0.006)
Log firm size	0.076***
	(0.003)
Constant	-1.291***
	(.008)
N	49,736
R^2	0.201

Sample of new lawyers entering firms of size 4+
Robust std. errors (in parens)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Change in firm rank

	(1)	(2)	(3)
Poached	-0.061***	-0.063***	-0.063***
	(0.001)	(0.001)	(0.001)
Displaced	0.037***	0.036***	0.035***
	(0.001)	(0.001)	(0.001)
Firm Rank		-0.069***	-0.092***
		(0.001)	(0.001)
LSQ			0.018***
			(0.000)
Ln Mkt Size			0.002***
			(0.000)
Exper.			0.000
			(0.000)
Age			-0.000
			(0.000)
Constant	-0.001***	0.041***	0.040***
	(0.000)	(0.000)	(0.004)
Year FE	NO	NO	YES
Obs.	314,984	314,984	313,683
R2	0.024	0.058	0.070

Lawyers currently in firms of size 4+, 1933-1960, aged 22-55, non-attriting
Sample includes lawyers who are retained as omitted category (neither poached nor displaced)
Omitted year is 1933
Std. errors in parenthesis * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Linear Probability Models of Future A Rating

	Obtains A Rating		
	(1)	(2)	(3)
Poached	-0.048***	-0.045***	-0.041***
	(0.006)	(0.006)	(0.005)
Displaced	-0.010	-0.012	-0.001
	(0.010)	(0.010)	(0.010)
Firm-rank 0-25	0.299***	0.310***	0.358***
	(0.004)	(0.004)	(0.024)
Firm-rank 25-50	0.359***	0.346***	0.401***
	(0.003)	(0.003)	(0.024)
Firm-rank 50-75	0.413***	0.385***	0.424***
	(0.003)	(0.003)	(0.024)
Firm-rank 75-100	0.455***	0.408***	0.441***
	(0.002)	(0.003)	(0.024)
LSQ		0.072***	0.084***
		(0.002)	(0.002)
Added Ctrls	NO	NO	YES
Obs.	119,007	119,007	119,007
R^2	0.406	0.410	0.448

Sample includes lawyers who remain in firms for 10+ years
Added ctrls include log mkt-size, year FE, and age
Robust std. errors in parenthesis * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: 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 6: Model Estimates

Parameter	**Estimate** (std. error)	**Description**
δ	0.931	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 Analytical derivations of the objects in Definition 1

Using the assumed functional form in Assumption 8, here I analytically derive optimal output under various degrees of information about the worker's talent, defined in Definition 1.

The full information output is

$$\begin{aligned} \bar{y}^{FIM}[z_1, z_2] = \mathbb{E}_{z|z_1 \le z \le z_2} y(z,z) = (\phi - \alpha)\mathbb{E}_{z|z_1 \le z \le z_2}\left[z^{\alpha+(\phi-\alpha)}\right] \\ = (\phi-\alpha)\frac{z_2^{\phi+1} - z_1^{\phi+1}}{(z_2 - z_1)(1+\phi)}. \end{aligned} \tag{13}$$

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 $z_2 - z_1$.

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

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

$$
\begin{aligned}
\mathbb{E}_{z|z_1\leq z\leq z_2} y(\theta^\star, z) &= (\phi-\alpha)\left(\mathbb{E}_{z|z_1\leq z\leq z_2} z^{(\phi-\alpha)}\right)^{\frac{\phi}{\phi-\alpha}} \\
&= (\phi-\alpha)\left(\frac{z_2^{1+(\phi-\alpha)}-z_1^{1+(\phi-\alpha)}}{(z_2-z_1)(1+(\phi-\alpha))}\right)^{\frac{\phi}{\phi-\alpha}}.
\end{aligned}
\tag{14}
$$

Appendix B 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. Suppose the worker has private information about z. Conditional on the resumé and incumbent firm, the worker's payoffs do not depend on her private information.

Proof of Lemma 3. Let the market's beliefs at time t and stage 1 be described by the resumé $[z_{1,t}, z_{2,t}]$, and suppose that this differs from the worker's own private information. The proof will be inductive. First, in stage 1 of the final period T, there is obviously no scope for the worker's private beliefs to influence the set of available poaching offers, because there is nothing the worker can do to change the information set available to the poaching firms. The value function is summarized by the poaching offers, so the claim holds in the time T subgame.

Now, suppose that we know that in all subgames with $k - 1$ or fewer remaining periods, the worker's expected payoff will be unaffected by her own beliefs, given her resumé, attachment status, and incumbent firm. Now consider the subgame with only k remaining periods. In stage one, poaching offers are made, and these do not depend on the worker's private information. In stage two, a counteroffer is made. The worker's private information determines the probability that the incumbent firm will attempt to keep her. For example, the resumé might equal $[0, 0.5]$, the worker might privately know herself to be between $[0.5, 1]$, and the incumbent's cutoff might be 0.5. In this extreme case, the worker believes she will be retained, while the market believes that she will separate.

However, recall that the incumbent firm's counteroffer always makes the worker indifferent between accepting it, and rejecting it and receiving the value associated with the downgraded resumé, $V(z_{1,t}, \zeta_t)$, where ζ_t is the incumbent firm's equilibrium cutoff rule. According to Assumption 3, if the worker in the previous example were retained, her resumé would update to $[0.5, 0.5]$, while if she separated, it would remain at $[0, 0.5]$. Thus, although the worker's private information changes the probability of receiving an attractive counteroffer, this alone does not change expected payoffs.

Thus, the worker can receive at least $V(z_{1,t}, \zeta_t)$ by rejecting the incumbent's offer and accepting the most attractive poaching offer. The only way she can receive more than this would be by making an off-equilibrium path offer selection, and having market beliefs update accordingly. However, any off-path choice would have to be equally attractive to a worker that had no private information, which cannot happen in equilibrium.

□

Lemma 5.

$$w_t^R(\theta, z_1, z_2, \zeta) = V_t(z_1, \zeta) - \delta\left((1-\lambda^D)V_{t+1}(\zeta, \zeta_{t+1}(\theta, \zeta, z_2)) + \lambda^D V_{t+1}(\zeta, z_2)\right). \tag{2}$$

Proposition 1. The *equilibrium cutoff rule* is

$$\zeta_t(\theta, z_1, z_2) = \sup_{\zeta \in [z_1, z_2]} \left\{ y(\theta, \zeta) + (1-\lambda^D)\delta\pi_{t+1}^I(\theta, \zeta, z_2, \zeta) - w_t^R(\theta, z_1, z_2, \zeta) > 0, \forall z > \zeta \right\}. \tag{3}$$

Proof of Proposition 1. Suppose we are in stage 2 of period t, where incumbent firm θ_t has announced a retention wage w_t^R, the worker with resumé $[z_{1,t}, z_{2,t}]$ has applied, and the incumbent is now choosing whether or not to retain her after learning that her true talent equals z. Clearly, the firm retains the worker if and only if the payoff is positive. The payoff must be strictly positive for any worker who is strictly more talented than the marginally retained worker type.

So far, I have explained the necessity of the inequality inside the brackets. Now, to understand why the equilibrium cutoff must be the largest ζ satisfying this necessary cost-benefit condition, one must appreciate that poaching firms have a first mover advantage. By making offers in stage 1 of the period, they can force coordination on *any* value of ζ satisfying the cost-benefit condition. Thus, if some smaller ζ were being played in equilibrium, this would imply the existence of some poaching firm that had passed up on the opportunity to poach a worker of type higher than ζ by making a higher poaching wage offer and forcing the incumbent firm to use a larger cutoff rule.

□

Proposition 2. 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 2. Consider the cutoff inequality:

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

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 3. There exists a range of values for δ for which there is strictly positive poaching 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.

Proof of Corollary 3. By choosing $\delta = 0$, we can guarantee that there will always be some degree of poaching. To see why, notice that in this special case, the worker always accepts the highest wage offer in stage 3. Thus, in any situation where $\theta > z_1$, the type z_1 would be willing to offer a poaching wage equaling at least $y(z_1, z_1)$ which is strictly more than what the incumbent would be willing to pay for the type z_1 worker, since $y(z_1, z_1) > y(\theta, z_1) \forall \theta \neq z_1$. Hence, there must be strictly positive turnover, as long as $\theta > z_1$.

To see why $\theta > z_1$ along the equilibrium path, imagine the wage offers facing an unattached worker. The best offer will always come from the firm that maximizes expected output, $\theta^{\max}(z_1, z_2) > z_1$. Thus, $\theta > z_1$ everywhere along the equilibrium path.

By combining this with Proposition 2, we now know that if δ is sufficiently low, poaching occurs everywhere along the equilibrium path. It is easy to demonstrate that the equilibrium cutoff rule must be a continuous function of δ.

to conclude that as long as δ is not too high, there will still be some turnover at certain points along the equilibrium path. □

Lemma 9.

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

Proposition 3 (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 3. 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

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

which implies

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

$\square$

Proof of Lemma 10. I will prove the result by induction on the number of remaining periods in the game. Set all equilibrium objects for period $T+1$ equal to 0 so that they are, trivially, homogeneous. Suppose that Π^I_{t+1}, π^I_{t+1}, and V_{t+1} are homogeneous of degree ϕ, while ζ_{t+1} is homogeneous of degree 1. Then I will show several implications for the time t equilibrium objects, starting with the fact that V_t is homogenous of degree ϕ. Examine Equation 6, which is rewritten below.

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

The first component, the initial wage, is clearly homogeneous of degree ϕ. The third component, the continuation value of a displaced worker, is also clearly homogeneous of degree ϕ. The second component is less obvious, so I will illustrate this below.

$$
V_{t+1}(\lambda z_1, \zeta_{t+1}(\lambda\theta, \lambda z_1, \lambda z_2)) = V_{t+1}(\lambda z_1, \lambda\zeta_{t+1}(\theta, z_1, z_2)) = \lambda^\phi V_{t+1}(z_1, \zeta).
$$

The first equality uses degree-one homogeneity of the future cutoff rule, and the second uses degree-ϕ homogeneity of the value function. Thus, the time-t poaching indirect utility is homogeneous of degree ϕ, and thus V_t, its maximized value across all θ, must also be homogeneous of degree ϕ, while its maximizer, $\theta_t(z_1, z_2)$, must be homogeneous of degree 1.

Under the same assumptions, I will now show that the time t hypothetical retention wage is homogeneous of degree ϕ. Referring back to Lemma 5, we have

$$
w^R_t(\theta, z_1, z_2, \zeta) = V_t(z_1, \zeta) - \delta\left((1-\lambda^D)V_{t+1}(\zeta, \zeta_{t+1}(\theta, \zeta, z_2)) + \lambda^D V_{t+1}(\zeta, z_2)\right). \quad (15)
$$

which is a weighted sum of ϕ-homogeneous functions and thus also ϕ-homogeneous. Next I'll show that the equilibrium cutoff rule is homogeneous of degree 1. Referring back to Equation 3, we have

$$\zeta_t(\theta, z_1, z_2) = \sup_{\zeta \in [z_1, z_2]} \left\{ y(\theta, \zeta) + (1 - \lambda^D)\delta \pi^I_{t+1}(\theta, \zeta, z_2, \zeta) - w^R_t(\theta, z_1, z_2, \zeta) > 0, \forall z > \zeta \right\}. \tag{16}$$

The left-hand-side of the inequality inside the supremum is homogeneous of degree 1 in $(\theta, z_1, z_2, \zeta)$. Thus, ζ solves the inequality for (θ, z_1, z_2) if and only if $\lambda\zeta$ solves the inequality for $(\lambda\theta, \lambda z_1, \lambda z_2)$, implying that the solution must be homogenous of degree one.

Next, I will show that Π^I_t and π^I_t are also homogeneous of degree ϕ. Referring back to Equation 4 and Equation 5, we have

$$\begin{aligned} \Pi^I_t(\theta, z_1, z_2) = & \frac{z_2 - \zeta}{z_2 - z_1} \left(\bar{y}(\theta, \zeta, z_2) + \delta(1 - \lambda^D)\Pi^I_{t+1}(\theta, \zeta, z_2) - w^R \right), \\ & \text{subject to } \zeta = \zeta_t(\theta, z_1, z_2), \\ & \quad \text{and } w^R = w^R_t(\theta, z_1, z_2), \end{aligned}$$

which is clearly homogeneous of degree ϕ based on the previous results, and

$$\begin{aligned} \pi^I_t(\theta, z_1, z_2, z) = & \mathbf{1}(z > \zeta) \left(\bar{y}(\theta, z) + \delta(1 - \lambda^D)\pi^I_{t+1}(\theta, \zeta, z_2, z) - w^R \right), \\ & \text{subject to } \zeta = \zeta_t(\theta, z_1, z_2), \\ & \quad \text{and } w^R = w^R_t(\theta, z_1, z_2), \end{aligned}$$

which is also clearly homogeneous of degree ϕ. The claim follows by induction.

□

Proof of Lemma 11. The first equation follows by applying Assumption 6 in order to substitute $\ln z_1$ in Corollary 4. The second equation is more involved. I assume that a lawyer obtains an A rating if and only if her talent is above some constant threshold Z_A. The assumption that the probability of achieving an A rating is strictly between 0 and 1 implies that z_A is in the interior of $[z_1, z_2]$. Combining this with Assumption 6 gives the following closed form relationship mapping the $\mathbf{x}$-conditional probability of receiving an A rating to the value $g(\mathbf{x})$.

$$
\begin{aligned}
p_A(\mathbf{x}) = Pr(z > z_A|\mathbf{x}) = \mathbb{E}_{z_1,z_2}\left[\frac{z_2 - z_A}{z_2 - z_1}|\mathbf{x}\right] &= \mathbb{E}_{z_1,z_2}\left[\frac{\gamma z_2 - \gamma z_A}{(\gamma-1)z_2}|\mathbf{x}\right] \\
&= \mathbb{E}_{z_1,z_2}\left[\frac{\gamma \exp\left(g(\mathbf{x})+\epsilon\right) - \gamma z_A}{(\gamma-1)\exp\left(g(\mathbf{x}_j)+\epsilon\right)}|\mathbf{x}\right] \\
&= \frac{\gamma}{\gamma-1}\frac{e^{g(\mathbf{x})} - z_A \mathbb{E}_{z_1,z_2}\left[e^{-\epsilon}|\mathbf{x}\right]}{e^{g(\mathbf{x})}} \\
&= \frac{\gamma}{\gamma-1}\left(1 - z_A \mathbb{E}\left[e^{-\epsilon}\right] e^{-g(\mathbf{x})}\right).
\end{aligned} \tag{17}
$$

Solving for $g(\mathbf{x})$, we have

$$
g(\mathbf{x}) = \ln\left(z_A \mathbb{E}\left[e^{-\epsilon}\right]\right) - \ln\left(1 - \frac{\gamma-1}{\gamma} p_A(\mathbf{x})\right).
$$

Plugging this in to the first equation, and letting the constant $\tilde{v}_0$ equal $\ln V(1,\gamma) + \phi \ln\left(z_A \mathbb{E}\left[e^{-\epsilon}\right]\right)$, we have

$$
\tilde{v}(\mathbf{x}) = \tilde{v}_0 - \phi \ln\left(1 - \frac{\gamma-1}{\gamma} p_A(\mathbf{x})\right). \tag{18}
$$

□

Proposition 4 (Identification of ϕ and γ)**.** 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

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

Proof of Proposition 4. Let $\tilde{\gamma} = \frac{\gamma-1}{\gamma}$. Taking derivatives of $\tilde{v}$ in Equation 18 yields

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

and

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

Square the first expression and divide it by the second to identify ϕ. Then plug the result for ϕ into the first expression and solve for γ to identify γ.

$\square$

Appendix C 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.[56]

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.[57]

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

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

[57]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.[58] 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.[59] Figure 7 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.[60]

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.

[58] 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.

[59] 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.

[60] 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 7.

$$r_{i,f,t} = \frac{1}{N_f} \sum_{j=1}^{N_t} \mathbf{1}\left(\widehat{LSQ}_{j,f,t} < \widehat{LSQ}_{i,f,t}\right) \tag{19}$$

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 7: Cardinal measures of law school quality

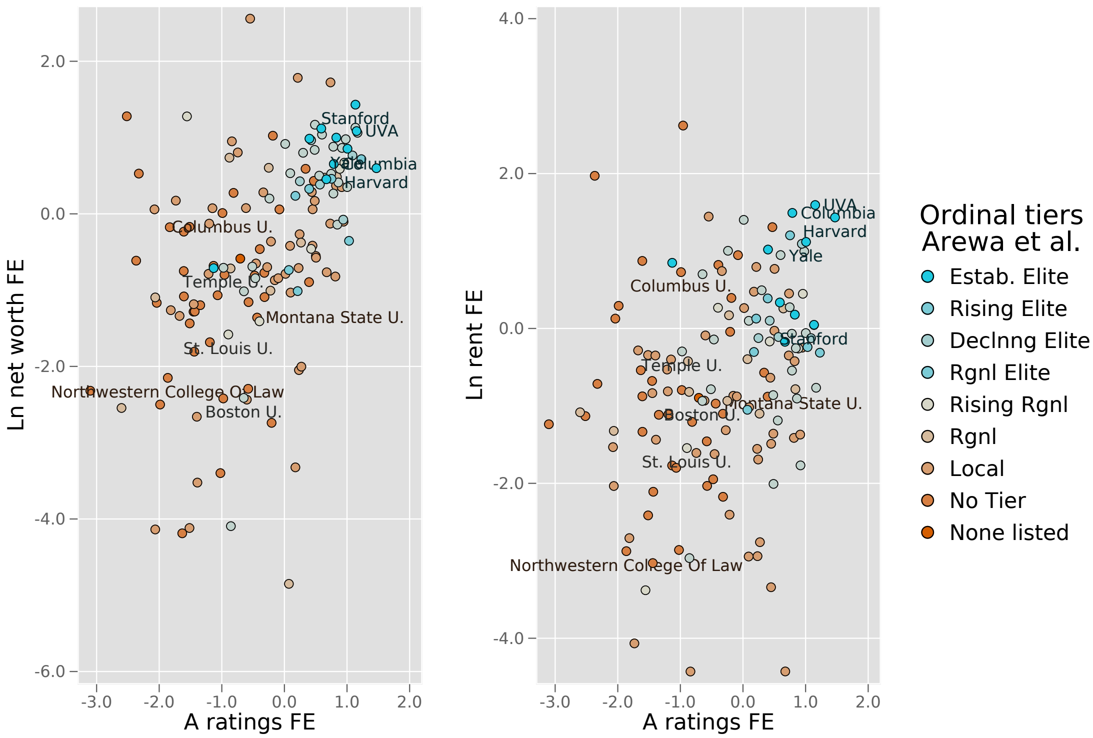

Table 7: Career success vs. firm rank

	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.

Fuchs, William and Andrzej Skrzypacz (2019), "Costs and benefits of dynamic trading in a lemons market." *Review of Economic Dynamics*, 33, 105 – 127, URL http://www.sciencedirect.com/science/article/pii/S1094202518303089. Fragmented Financial Markets.

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.