Title
Matching, Reallocation, and Retention in Labor Markets

Author
Partow, Rustin John

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Matching, Reallocation, and Retention in Labor Markets

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Rustin John Partow

2020

ABSTRACT OF THE DISSERTATION

Matching, Reallocation, and Retention in Labor Markets

by

Rustin John Partow
Doctor of Philosophy in Economics
University of California, Los Angeles, 2020
Professor Moshe Buchinsky, Co-Chair
Professor Maurizio Mazzocco, Co-Chair

This dissertation is devoted to studying how workers initially match with firms, and are subsequently retained or reallocated over time. The first two chapters—one theoretical, the other empirical—specifically ask how asymmetric employer information distorts career placement and efficiency in a dynamic setting. The final chapter studies optimal dynamic compensation policy from the standpoint of a single employer attempting to maximize retention at minimum cost.

In Chapter 1, I develop a new framework to investigate the efficiency costs of asymmetric information in the labor market. I consider a setting where, similar to Greenwald (1986), employers have an inside advantage in learning their employees' abilities. I then add complementarity between firms' heterogeneous technologies and the abilities of their workers in order to study how asymmetric information can disrupt the efficient assignment of workers to firms. Workers' abilities are initially hidden and become privately observed by their employers. Incumbent firms retain high-ability workers. Low-ability workers separate to uninformed but comparatively advantaged outsiders where the returns to ability are low. Relative to a static optimum, new hires over-place into inefficiently high-type firms, and then become under-placed as they accumulate tenure. These placement distortions present an added source of inefficiency relative to a symmetric learning environment. I derive a vari-

ety of testable implications of the model, and prove a non-parametric method for identifying the surplus function.

In Chapter 2, I present evidence on how adverse selection and production complementarities interact to produce distinct patterns in reallocation. I build a new data-set on the US market for lawyers by linking together the Martindale-Hubbell professional directories from 1930-1963. I show evidence that lawyers who separate from surviving firms are adversely selected, and move to firms where their peers have lower average ability. Meanwhile, lawyers who separate after their firm exits move to firms with higher-ability peers, but are *not* positively selected compared to similar lawyers who are retained. These results provide evidence to support the asymmetric learning model of Chapter 1.

In Chapter 3, which is co-authored with Moshe Buchinsky and John de Figueiredo, we examine the cost-effectiveness of compensation policy in the federal government. We estimate a dynamic retention model using the federal government's personnel data, exploiting exogenous pay variation caused by the 1990s civil service pay reform known as the Federal Employees Pay Comparability Act (FEPCA). We find that the elasticity of retention to pay is typically around 25% for the workers in our sample. The model can be combined with assumptions about government hiring in order to make long-run out-of-sample forecasts of payroll costs, turnover, and workforce composition under alternative compensation policies.

The dissertation of Rustin John Partow is approved.

Moritz Meyer-ter-Vehn

Dora Luisa Costa

Christopher L. Erickson

Maurizio Mazzocco, Committee Co-Chair

Moshe Buchinsky, Committee Co-Chair

University of California, Los Angeles

2020

To my mother. . .

for filling my head with dreams,

and father. . .

for filling my dreams with inquiry.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am most of all indebted to Maurizio Mazzocco and Moshe Buchinsky, who co-chaired my dissertation committee. I would not be the economist I am today without their careful guidance, and I will do my best to pay the debt forward. Moritz Meyer-ter-Vehn was a tremendous help in developing the theory in my first chapter, and the applied work in my second chapter would not have been possible without Dora Costa's advice and the research resources she has marshaled on behalf of the economic history group. I also thank Chris Erickson for his unique perspectives on my research.

I am grateful for the training I received at UCLA, and I want to extend particular thanks to John Asker and Simon Board for providing long-lasting advice on how generally to be a good economist, especially when it came to communicating ideas effectively. My research also benefited from conversations with: Carolina Arteaga, Victoria Barone, El Hadi Caoui, Sepehr Ekbatani, Alexandre Fon, Jonathan Gu, Tomas Guanziroli, Hugo Hopenhayn, Ariadna Jou Fuya, Keyoung Lee, Vladimir Pecheu, Bernardo Silveira, Sumit Shinde, Renato Zaterka, Diego Zuñiga, and seminar participants from the University of Minnesota, University of Virginia, UCLA.

The best part of this experience was pitching research to peers. I want to thank Daehyun Kim for being a wonderful coauthor—it has been a true pleasure to work with him. I want to thank those friends who never tired of talking economics, especially Brett McCully, Sumit Shinde, and Vitaly Titov. And finally, I want to thank the friends with whom I built lasting memories of our time in LA. I thank the California Center for Population Research, and especially Ana Ramirez, for creating a friendly and hospitable research environment in which I spent five happy years. Finally, I gratefully acknowledge financial support from the All-UC Group in Economics and the Center for Economic History.

VITA

2007 – 2011	B.S. in Economics, Math at George Washington University.
2011 – 2013	Consultant at Bates White LLC.
2013 – 2014	M.A. in Economics at University of Texas at Austin.
2014 – 2020	C.Phil. in Economics at University of California at Los Angeles (UCLA).
2015-2019	Teaching Assistant at UCLA.
2020	Dissertation Year Fellowship, Graduate Division, UCLA.
2017	All-UC Group Student Research Grant, 2017
2017	Center for Economic History Student Grant, 2017
2017	Graduate Summer Research Mentorship, Graduate Divison, UCLA
2016-2017	Graduate Research Mentorship, Graduate Divison, UCLA
2016	Proctor of the Year, Economics Department, UCLA

CHAPTER 1

Adverse Selection and Complementarity in the Labor Market

1.1 Introduction

Many labor markets feature firm-worker complementarity, in which it is efficient for some firms to be matched to higher ability workers than others. However, when workers enter the labor market, their ability is often highly uncertain. By monitoring their new hires at work, employers are likely to resolve much more of this uncertainty than the rest of the market. In such an environment, how efficient are the initial placements? And how much, if at all, do the private observations of employers eventually benefit society through reallocation?

To answer this question, I develop a new modeling framework that combines firm-worker complementarity and asymmetric employer learning. Each firm's comparative advantage is to hire a worker whose ability matches the distinct return to ability inherent to its technology. If workers' abilities were immediately known, then the model would feature immediate perfectly assortative matching. Instead, each worker enters the labor market with a noisy signal, resulting in an imperfect initial match. Her new employer privately learns her ability, and competes against ill-informed outside firms to retain her. A classic lemons problem ensues (Akerlof, 1970), where high-ability workers are retained and only low-ability workers are relinquished to outsiders.

Returns to assortatively matching firm and worker types ensure that some endogenous separations occurs in spite of the adverse selection problem.[1] Outside offers come from

[1]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

lower-type firms whose technology gives them a comparative advantage in hiring low ability workers.

Workers are publicly revealed to be above or below a threshold based on whether or not they are retained, leading to increasingly precise Bayesian inference about their types. Hence, the large initial inefficiencies of the lemons problem are rectified over time through reallocation and information leakage. However, this reallocation is distorted relative to a symmetric learning setting. Retained workers eventually become under-placed at the current firm, and are inefficiently restrained by the adverse selection problem from moving up to a higher-type firm. Because of the future prospect of under-placement, new matches feature *over-placement*. The simple cutoff dynamics of beliefs is a major advantage of the model's equilibrium, ensuring that the analysis remains tractable as the time-horizon extends to infinity—a rare feat for an asymmetric learning model.

In addition to predicting that endogeneously separating workers move to lower-type firms, the model makes the stark prediction that workers who are displaced by exogenous firm exits will mvoe to *higher*-type firms, due to the aforementioned under-placement phenomenon. The model provides a tractable quantitative framework for evaluating allocational efficiency and assessing potential labor market reforms.

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 1.2 presents the model, solves it and derives a set of testable implications as well as normative predictions. Section 1.4 explains how to identify the model using employee-employer matched data. Section 1.5 describes several model-based tests. Section 1.6 concludes.

1.1.1 Related Literature

Learning about ability. I build on the employer learning literature. The idea that asymmetric information between employers distorts mobility and impedes the efficient assignment

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.

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 confer different degrees of status when their names appear on resumés.[2,3]

One of the main predictions from the literature on asymmetric learning in the labor market has been that job assignments within firms will signal private information, and that firms will tend to under-promote relative to a first-best setting in order to prevent the resulting loss of their informational rents, with Waldman (1984) being the first to point this out, and DeVaro and Waldman (2012) establishing empirical evidence that promotions indeed carry signaling content. There has been comparatively little work on the extent to which asymmetric learning distorts assignment *to* firms, rather than within firms.

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 predicts that endogenously separating workers move to higher type firms, while I predict the opposite.[4]

[2]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.

[3]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.

[4]There are some other important differences. My model admits a wider range of firm and worker

My model features dynamic updating of beliefs about each worker's ability through an evolving posterior, 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 to admit an arbitrarily long time horizon without losing tractability, 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.[5] My model makes a similar but not identical prediction. According to my model, wages should become less correlated with ex ant characteristics like education and race and more correlated with true ability over time. However, the relationship between true ability and wages should be entirely mediated through the worker's public job history.

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

heterogeneity in order to make the model relevant to empirical work.

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

by abstracting from how information influences the quality of firm-worker matches. The main contribution of my model is to present a framework that economists can use to assess the ex-ante social benefits of having more precise information about ability.

1.2 The Model

Endowments and technology. There are infinitely many firms indexed by type $\theta \in R^+$ that have the option of either participating in the 'inside market', or participating in an 'outside market' with known period profits $C(\theta)$, which are increasing in θ. The inside market is initially endowed with a stock of workers with randomly distributed ability z which is normalized to be uniformly distributed between two endpoints, $z_{1,1}$ and $z_{2,1}$.[6] Firm-worker pairs produce output according to $f(\theta, z)$, which is increasing in z and θ.[7]

Structure of the game. The model is a repeated game played between workers and firms. Time $t = 1, 2, \ldots$ is discrete and infinite. Each worker begins a period with an *attachment status* θ denoting the type of her current employer, or *incumbent* firm, with $\theta = \emptyset$ implying that she is initially unattached. The incumbent privately observes the worker's ability z, while the entrepreneurs and the worker simply know the public history. All actions are public.

Offer-making. Each worker will be put up for a bidding in which firms have the opportunity to make competing spot wage offers to her. The attached workers are put up for bidding first. Vacant firms can make the worker an outside offer, and then the incumbent can make a counteroffer. After the incumbent makes the counteroffer, it observes a mean-0 i.i.d. shock ε to the payoff of retaining the worker, and can choose to adjust its offer by the amount ε. The worker then chooses any of the previously made offers, or quits to become unattached.

[6] Hence, a worker's z measures their percentile in the ability distribution. The assumption of a one-dimensional, time-invariant talent or "individual competency" (Postel-Vinay and Robin, 2002) parameter is fairly standard in the literature.

[7] My preferred interpretation of the outside market is that it is a market that produces the same goods and services but where workers' abilities are known with certainty.

Every offer and counteroffer bears an infinitesimally small writing cost. The bidding then moves to unattached workers (including any who recently quit). This time, only the vacant firms have the opportunity to make offers, and the worker chooses her preferred offer. If she gets zero offers, the worker exits the industry.[8]

Period payoffs. At the end of a period, the workers payoff is the accepted wage. Each winning firm loses the offered wage and gains output. Each firm type has a fixed opportunity cost of operation, $C(\theta)$, so I will refer to its *surplus* as $s(\theta, z) = y(\theta, z) - C(\theta)$ and I assume that a winning firm's payoff is $s(\theta, z)$ (or $s(\theta, z) + \varepsilon$, for incumbents) minus the accepted wage. Losing firms obtain zero payoffs.

Death and transitions. If a winning firm and its worker both survive, then they enter next period attached. Workers live to be at most T periods old, and along the way have a chance of dying (exiting the sector permanently) at exogenous rate $1 - \delta$. A firm dies if it makes an offer that goes unaccepted, or if its worker dies.[9] Otherwise, firms can also die exogenously at rate $1 - \lambda$. If the firm had previously employed a worker, then she enters next period unattached.

Definition of an equilibrium. An equilibrium is a collection of (1) beliefs mapping the history of play into a probability distribution over each worker's ability, and (2) strategies mapping histories of play into wage-offers, counter-offers, adjusted counteroffers, quit decisions, and offer-acceptances. I will search for a Perfect Bayesian Equilibrium (PBE). The Perfect Bayesian restriction requires that agents' beliefs be consistent with Bayes' rule, wherever it applies, and that the decision rules be sequentially rational according to those beliefs. I make the following additional refinements.

---

[8]The assumption that incumbents respond sequentially to poaching offers is the most common modeling choice among recent asymmetric employer learning papers, but see Li (2013) for an important counter-example. All-else equal, alternative timings are likely to change the amount of turnover predicted by the model. My identification strategy in Section 1.4 will be robust to misspecification of this timing, but counterfactual estimates will not be. I explain how to test the timing assumption in Section 1.5.

[9]This assumption is similar to Assumption 9 in (Bernhardt, 1995, p. 319), and is made to ensure that the relevant history remains tractable. However, it is possible to show that in equilibrium firms would never re-match with a worker from whom they had previously endogenously separated.

Equilibrium refinement. The shock ε is equal to zero. The equilibrium must be the limit of a sequence of equilibria corresponding to a sequence of random variables, $\{\varepsilon\}_{k=1}^{\infty}$, in which each ε_k has continuous support covering the real line. For most of the analysis, it will suffice to ignore ε. But this refinement helps ensure that off-path beliefs about quitting workers are plausible. The refinement is similar in spirit to the Trembling Hand (Selten, 1975) and Sequential (Kreps and Wilson, 1982) refinements in the literature.

For this next assumption, note that a multivariate function $f(\mathbf{v})$ is homogeneous of degree (HD-) α if and only if $f(t\mathbf{v}) = t^{\alpha} f(\mathbf{v})$ for all $t \geq 0$.

Assumption 1 (Homogeneity)**.** The surplus function $s(\theta, z)$ is HD-ϕ in θ and z.

This assumption is made for computational tractability as the homogeneity properties will carry through to other equilibrium functions. It is not necessary for most of the qualitative results.

Assumption 2 (Concavity and one-to-one optimal matching)**.** $s(\theta, z)$ is strictly concave in θ and, fixing z, is maximized at $s^{\star}(z) = s(\theta^{\star}(z), z) > 0$. $\theta^{\star}(z)$ is either strictly increasing or decreasing in z.

This assumption implies that in a full-information setting, all workers are matched to a unique firm type, and that the farther a worker departs from the optimal matching, the less surplus she creates.

1.3 Solving the Model

1.3.1 Key Features of the Equilibrium.

In what follows, let $t \in 1, 2, ..., T$ denote a worker's age in the current period. Let θ_t denote the worker's attachment status at the start of the period. I will be particularly interested in matches that are not 'grossly' misplaced in the following sense.

Definition 1. Consider a worker attached to θ_t with posterior ability between $z_{1,t}$ and $z_{2,t}$. A worker is misplaced if θ_t's full-information match is below $z_{1,t}$ (or above $z_{2,t}$).

The following claim, which I will prove later on, is the core feature of the equilibrium.

Claim 1 (Cutoff rule retention and quits)**.** Fix (t, θ_t) and assume that the worker is not misplaced. In equilibrium, θ_t makes a counter-offer if and only if it privately knows that z is larger than some cutoff Z which depends on the public history, and which occurs with probability strictly between 0 and 1. The amount of the counteroffer, if made, is constant. The counteroffer is never adjusted, and the worker never quits. Off-path, if the worker quits, she is believed to have ability equaling the cutoff Z.

This feature says nothing about misplacement. In fact, such matches can have multiple equilibrium outcomes, but I will show that regardless of which outcome is played, such matches will never arise endogenously and can therefore be safely ignored from the analysis of the equilibrium path. The next two corollaries immediately follow, describing how posterior beliefs update after observed counteroffers.

Corollary 1. Suppose Claim 1 holds and the worker is not misplaced. Then the posterior distribution $\mathbf{z}_t$ of a worker's ability is $z \sim \mathcal{U}(z_{1,t}, z_{2,t})$, where the end-points evolve according to the following recursive rule:

1. Begin the worker's life at $z_{1,1} = 0$ and $z_{2,1} = 1$.
2. After a counter-offer is made, raise z_1 to equal the cutoff. If the worker subsequently quits, lower z_2 to equal the cutoff.
3. After a counteroffer is withheld, lower z_2 to equal the cutoff
4. Otherwise, both cutoffs remain constant.

Claim 2 (Markov Perfect Equilibrium)**.** Let $(z_{1,t}, z_{2,t})$ characterize a worker's posterior ability. Suppose that she is attached to incumbent $\theta_t \in (z_{1,t}, z_{2t})$. Her expected equilibrium earnings at the beginning of the period can be described by some function $v_t(\theta_t, z_{1,t}, z_{2,t})$. If she is unattached, her equilibrium earnings can be described by some function $V_t(z_{1,t}, z_{2,t})$ and the vacant firm that she matches to as $\theta_t(z_{1,t}, z_{2,t})$. The amount of the incumbent's counteroffer, if one is made, can be described by some function $w^R(z_{1,t}, z_{2,t})$. The incumbent's payoff is, conditional on true ability z, described by some function $\Pi_t(\theta_t, z_{1,t}, z_{2,t}, z)$. All equilibrium strategies and beliefs are functions of $(t, \theta_t, z_{1,t}, z_{2,t})$ which evolve in equilibrium according to a first-order Markov process.

Henceforth, refer to $\upsilon_t(\theta_t, z_{1,t}, z_{2,t})$ as the attached value function, $\theta_t(z_{1,t}, z_{2,t})$ as the equilibrium assignment function, $V_t(z_{1,t}, z_{2,t})$ as the unattached value function, $Z_t(\theta_t, z_{1,t}, z_{2,t})$ as the cutoff rule function, $w_t^R(z_{1,t}, z_{2,t})$ as the retention wage, and $\pi_t(\theta_t, z_{1,t}, z_{2,t}, z)$ as the ex-post incumbent profit function. Also, define the *ex-ante* incumbent profit function as

$$\Pi_t(\theta_t, z_{1,t}, z_{2,t}) \equiv \mathbb{E}_z\left(\pi_t(\theta_t, z_{1,t}, z_{2,t}, z) | z_{1,t} \le z \le z_{2,t}\right).$$

All of the above functions are defined to equal 0 if $t > T$. In equilibrium, if a vacant firm successfully wins a worker, it must earn exactly zero expected profits from doing so. Otherwise, a vacant firm of the exact same type would be willing to deviate by making a strictly higher offer, which the worker would be willing to accept. The next lemma follows by definition of $\Pi_{t+1}(\theta^v, z_{1,t}, z_{2,t})$.

Lemma 1 (New-hire wage)**.** Suppose Claim 1 and Claim 2 hold. Suppose that a vacant firm θ^v successfully hires an age-t worker in equilibrium who has posterior ability $(z_{1,t}, z_{2,t})$ when she accepts the offer. Then it pays her a new-hire wage w^H satisfying

$$w_t^H(\theta^v, z_{1,t}, z_{2,t}) = \mathbb{E}_z\left(s(\theta^v, z) | z_{1,t} \le z \le z_{2,t}\right) + \delta\lambda\Pi_{t+1}(\theta^v, z_{1,t}, z_{2,t}). \quad (1.1)$$

The value of joining the vacant firm is the new-hire wage above, plus the continuation value associated with becoming attached to it next period, or (if it exits) becoming unattached. The winning vacant firm is the one that provides the highest overall value. Moreover, in the case where z is known with certainty $(z_{1,t} = z_{2,t})$, the worker's full-information match is willing to pay her full-information payoff $s^\star(z_{1,t})$ in every remaining period of her life. This gives a recursive result for the unattached value function.

Lemma 2 (Unattached value function)**.** Suppose Claim 1 and Claim 2 hold, and the worker is not misplaced. Then

$$\begin{aligned} V_t\left(z_{1,t}, z_{2,t}\right) =& \max_{\theta^v} \tilde{V}_t(\theta^v, z_{1,t}, z_{2,t}) \\ \text{where } V_t(\theta^v, z_{1,t}, z_{2,t}) =& w_t^H(\theta^v, z_{1,t}, z_{2,t}) + \delta\left(\lambda v_{t+1}\left(\theta^v, z_{1,t}, z_{2,t}\right) + (1-\lambda) V_{t+1}\left(z_{1,t}, z_{2,t}\right)\right). \end{aligned} \tag{1.2}$$

Moreover, if $z_{1,t} = z_{2,t}$, then

$$V_t\left(z_{1,t}, z_{2,t}\right) = S_t\left(z_{1,t}\right) \equiv \sum_{t=1}^{T} \delta^t s(\theta^\star(z), z).$$

Lemma 3 (Retention wage)**.** Suppose Claim 1 and Claim 2 hold and consider an attached worker whose incumbent uses cutoff rule Z. Then if the worker is retained, her continuation value is $S_t(Z)$. The incumbent pays a retention wage of

$$\begin{aligned} w_t^R(\theta_t, z_{1,t}, z_{2,t}) = S_t(Z) + \delta\left(\lambda v_{t+1}(\theta_t, Z, z_{2,t}) + (1-\lambda) V_{t+1}(Z, z_{2,t})\right), \\ \text{where } Z = Z_t(\theta_t, z_{1,t}, z_{2,t}). \end{aligned} \tag{1.3}$$

See proof on page 74.

Lemma 4 (Cutoff rule)**.** Suppose Claim 1 holds. Fix $(t, \theta_t, z_{1,t}, z_{2,t})$ at the start of the period and assume that θ_t is the full information match for some $z \in \operatorname{supp}\left(z | z_{1,t} < z \leq z_{2,t}\right)$. Then the cutoff $Z = Z_t(\theta_t, z_{1,t}, z_{2,t})$ solves

$$s(\theta_t, Z) = w_t^R(\theta_t, z_{1,t}, z_{2,t}). \tag{1.4}$$

Proof. Claim 1 says that as long as $t < T$ and the incumbent is above the full-information match for $z_{1,t}$, then the worker always has a chance of being retained at a cutoff that *remains* below the incumbent's full-information match (until the final period). Consequently, a worker who is marginally retained at age t is guaranteed *not* to be retained in at age $t+1$. Thus, the incumbent's benefit of retaining the marginal type $Z = Z_t(\theta_t, z_{1,t}, z_{2,t})$ worker is the surplus $s(\theta_t, Z)$ and nothing more. The cost is the retention wage. Hence, the marginally retained worker type is paid her wage. $\square$

Lemma 5 (Attached value function)**.** Suppose Claim 1 and Claim 2 hold and fix $(\theta_t, z_{1,t}, z_{2,t})$. Let $Z = Z_t(\theta_t, z_{1,t}, z_{2,t})$. If the worker is retained—i.e., receives a counteroffer—(with probability p^R), her continuation value is $S_t(Z)$. If the worker is not retained, she receives continuation value $V(z_{1,t}, Z)$. Hence,

$$\begin{aligned}\upsilon_t\left(\theta_t, z_{1,t}, z_{2,t}\right) &= \left(1-p^R\right)V(z_{1,1}, Z) + p^R S_t(Z),\\ \text{where } p^R &= \frac{1-Z}{z_{2,t}-F(z_{1,t})},\\ \text{and } Z &= Z_t(\theta_1, z_{1,1}, z_{2,1}).\end{aligned} \tag{1.5}$$

Proof. Vacant firms who make the initial offers correctly anticipate that they only win when the incumbent firm doesn't make a counteroffer, in which case the worker's posterior is $(z_{1,t}, Z)$. Thus, the vacant firms bid on the worker as if she was unattached of type $(z_{1,t}, Z)$. Hence, if she doesn't get a counteroffer (with probability $1-p^R$), the worker's continuation value is $V_t(z_{1,t}, Z)$.

With probability p^R, the worker is retained at cutoff Z, and from Lemma 3, we know that she receives $S_t(Z)$. $\square$

Lemma 6 (Profit)**.** Suppose Claim 1 and Claim 2 hold. Then the ex-post profit function is

$$\begin{aligned}\pi_t(\theta_t, z_{1,t}, z_{2,t}, z) &= \begin{cases} 0 & \text{if } z \le Z \\ s(\theta_1, z) - w^R + \lambda\delta\pi_{t+1}(\theta_t, Z, z_{2,t}, z) & \text{if } z > Z, \end{cases}\\ \text{where } w^R &= w_t^R(\theta_t, z_{1,t}, z_{2,t})\\ \text{and } Z &= Z_t(\theta_t, z_{1,t}, z_{2,t}).\end{aligned} \tag{1.6}$$

Proof. The current ex-post profit is the surplus $s(\theta_t, z)$ minus the retention wage. Then, with probability $\delta\lambda$, the worker and firm both survive, and the incumbent receives the ex-post incumbent profit associated with next period. $\square$

The ex-post profit function follows immediately from integrating the ex-post profit function with respect to the posterior distribution of ability.

Corollary 2. The ex-ante profit function is give

$$\begin{aligned}
\Pi_t(\theta_t, z_{1,t}, z_{2,t}) = p^R \left(\mathbb{E}_z \left(s(\theta_1, z)|z_{1,t} < z \leq z_{2,t}\right) - w^R + \lambda\delta\Pi_{t+1}(\theta_t, Z, z_{2,t})\right), \\
\text{where } w^R = w_t^R(\theta_t, z_{1,t}, z_{2,t}), \\
p^R = \frac{1 - Z}{z_{2,t} - F(z_{1,t})}, \\
\text{and } Z = Z_t(\theta_t, z_{1,t}, z_{2,t}).
\end{aligned} \tag{1.7}$$

Proposition 1. Claim 1 and Claim 2 hold.

See proof on page 75.

Proposition 2. The equilibrium functions have the following properties.

1. The cutoff rule is HD-ϕ in $(\theta_t, z_{1,t}, z_{2,t})$ and increasing in t
2. The retention probability is HD-0 in $(\theta_t, z_{1,t}, z_{2,t})$ and decreasing in t
3. The unattached value function is HD-ϕ in $(\theta_t, z_{1,t}, z_{2,t})$, convex in $(z_{1,t}, z_{2,t})$, and decreasing in t
4. The equilibrium assignment function is HD-1 in $(\theta_t, z_{1,t}, z_{2,t})$ and decreasing in t
5. The attached value function is HD-ϕ in $(\theta_t, z_{1,t}, z_{2,t})$ and increasing in t
6. The ex-ante incumbent profit function is HD-ϕ in $(\theta_t, z_{1,t}, z_{2,t})$

See proof on page 76.

Recall that $\theta_t(z_{1,t}, z_{2,t})$ is the firm type that maximizes $\tilde{V}_t(\theta^v, z_{1,t}, z_{2,t})$, which I defined in Lemma 2. By analyzing the function $\tilde{V}_t(\theta^v, z_{1,t}, z_{2,t})$, I can derive some comparative statics relating posteriors to equilibrium matching.

Proposition 3 (Assortative matching on posterior ability)**.** $\theta_t(z_{1,t}, z_{2,t})$ is

1. HD-1;
2. $\leq \theta^\star(z_{2,1})$, with equality if and only if $z_{1,t} = z_{2,t}$;

3. $\geq \theta^\star(z_{1,1})$, with equality if and only if $z_{1,t} = z_{2,t}$;

4. Increasing in $z_{1,t}$ and $z_{2,t}$.

5. Decreasing in t

Finally, an analytically useful feature of the equilibrium is that the functions become stationary as the remaining time-horizon extends to infinity.

Proposition 4 (Infinite-horizon Markov equilibrium)**.** The equilibrium functions converge uniformly to stationary functions.

Proof. As T goes to infinity, the equilibrium functions converge uniformly to stationary functions. I proved that each equilibrium function is monotonic in time. It is also trivial to show that each function belongs to a compact set. Thus, by the monotone convergence theorem, each equilibrium function converges pointwise. Dini's theorem (Rudin, 1976, p. 150) then says that pointwise convergence of a monotonic sequence of continuous functions along a compact set implies uniform convergence. Thus, each equilibrium function converges uniformly. Since there are a finite number of equilibrium functions, the collection also converges uniformly. □

1.3.2 How to Solve for the Equilibrium

I will now explain how one can computationally solve the model, which is at a minimum necessary to produce counterfactuals and which could be also be used to estimate the model via indirect inference. To solve the model analytically, one would apply backward recursion to the equilibrium objects using the formulas in Section 1.2.

Setting T to be some arbitrary number, we will move back in worker's age t. Given knowledge of the age $t+1$ objects, we can recover the age t objects using the following recursive procedure (the order matters, as some of the time t objects rely on other time t objects):

1. Calculate the unattached value function using Lemma 2.

2. Calculate the retention wage using Lemma 3.

3. Calculate the equilibrium cutoff rule using Lemma 4.

4. Calculate the ex-ante incumbent profit function using Corollary 2 (calculating the ex-post incumbent profit function is unnecessary).

5. Calculate the attached value function using Lemma 5.

Computational methods. Because these functions lose analytical tractability after a few iterations, it is essential to use numerical methods to solve the model for doing quantitative exercises like simulation or structural estimation. To that end, I will enumerate several dimension reduction techniques that can be used when solving the model.

First, every equilibrium function features homogeneity. Thus, it is only necessary to solving for the equilibrium functions in the cases where $z_2 = 1$. Then, one can always recover those functions cases where $z_2 \neq 1$ by applying the corresponding degree of homogeneity in Proposition 2. For example, $V(t, z_1, z_2) = z_2^{\phi} V(t, \frac{z_1}{z_2}, 1)$.

Second, workers are never 'misplaced' in equilibrium, which means that one only needs to solve for the equilibrium functions in the domain where $\theta_t \in [z_{1,t}, z_{2,t}]$.

Third, given that a worker is not misplaced, there is always an interior solution to the cutoff equation in Equation 1.4. Given that an interior solution exists, it clearly does not depend on $z_{1,t}$—only on θ_t and $z_{2,t}$. Combining this with the earlier points, one needs to solve for the cutoff rules when $z_{1,t} = 0$ and $z_{2,t} = 1$.

1.3.3 Model Predictions

The model makes several interesting and empirically testable predictions about worker-firm dynamics.

Initial over-placement, eventual under-placement. An unattached worker in the final period T places with the firm that maximized static expected surplus. The final result of Proposition 3 says that young workers with small t place higher than older workers where

$t = T$. The final period T is a static constrained optimum benchmark where an unattached worker simply matches with the firm that maximizes expected surplus. Thus, young workers *over-place* relative to what is dictated by static efficiency. On the other hand, workers who enter period T attached are always *under-placed*. Since θ_t is the full-information match for the marginally retained type, it must be strictly below the optimal match for the full distribution of retained types. The optimal strategy for maximizing expected output, and therefore earnings, is to compensate for the tendency of retention to lead to under-placement by over-placing initially. As the number of remaining periods increases, so does the risk of eventual under-placement, causing equilibrium assignment to decrease in t.

These phenomena are added sources of inefficiency relative to a standard symmetric learning setting, where matches should maximize expected surplus based on the available information, as imperfect as it may be.[10]

Cutoff rules are ex-ante inefficient. Imagine if vacant firms could commit in advance to their sequences of cutoff rules and wages. Could they present an unattached worker with a better offer than what she receives in equilibrium? Again, the answer is yes. The equilibrium unattached value function is equal to the worker's net present expected surplus, so to answer this question we merely need to consider whether a marginal adjustment to a cutoff rule at some point during an employment spell results in strictly higher expected surplus.

Consider changing the cutoff rule Z in state $(t, \theta_t, z_{1,t}, z_{2,t})$ while leaving the rest of the equilibrium unchanged. Define $N_t(z)$ to denote the worker's total expected surplus along the equilibrium path if she is unattached in period t and has true ability z. Define $R_t(z)$ to denote the worker's total expected surplus along the equilibrium path if, instead, she is retained at the incumbent in period t. The worker's expected surplus at the start of the period is

[10]It is also possible to rationalize over-placement or under-placement by adding mobility costs and assuming a particular surplus function that results in the same asymmetric mobility predicted by this model. It is also possible to rationalize similar distortions if learning is endogenous to firm type, but in this latter case we should see no distortions in the final period $t = T$.

$$\tilde{S}_t(Z) = \frac{1}{z_{2,t} - z_{1,t}} \int_{z_{1,t}}^{z} N_t(z)dz + \frac{1}{z_{2,t} - z_{1,t}} \int_{Z}^{z_{2,t}} R_t(z)dz,$$

The derivative and second derivative of $\tilde{S}_t(Z)$ are

$$\tilde{S}_t'(Z) = \frac{1}{z_{2,t} - z_{1,t}} \left(N_t(Z) - R_t(Z)\right),$$

$$\tilde{S}_t''(Z) = \frac{1}{z_{2,t} - z_{1,t}} \left(N_t'(Z) - R_t'(Z)\right).$$

Conditional on the worker's true ability, z, the sequence of firms she is matched to in each future period is deterministic. Moreover, by the envelope theorem, we can treat this sequence as fixed when calculating the effects of marginal changes to the time t cutoff rule on future surplus. By Proposition 3, the type Z worker will be matched with strictly worse firms at all future points in time if separates at time t than if she is retained at time t. Since higher-type firms have higher returns to ability, $N_t'(Z) - R_t'(Z)$ is therefore negative, and we can conclude that $\tilde{S}_t(Z)$ is strictly concave. We can also verify that $\tilde{S}_t'(z_{1,t}) > 0$ as long as $\theta_t > \theta^\star(z_{1,t})$, in which case the separating worker would be matched to her full-information match and produce strictly more surplus than if she has to be retained at the incumbent for a period.

Hence, if the optimal cutoff is below $z_{2,t}$, then it must satisfy the first-order condition $N_t(Z) = R_t(Z)$. I.e., the marginal worker must generate equal net-present surplus streams if she is retained or if she separates. Compare this to the equilibrium cutoff condition:

$$s(\theta_t, Z) = S_t(Z) - \delta\left((1-\lambda)V_{t+1}(Z, z_{2,t}) + \lambda S_{t+1}(Z')\right),$$

where Z' is the future period's cutoff. In the case where $t = T$, the marginally retained worker is the incumbent's full-information match, which immediately implies that $N_t(Z) < R_t(Z)$. Hence, committing to a higher T cutoff rule could increase expected surplus, ceteris paribus. On the other hand, as the number of remaining periods goes to infinity and δ goes to 1, the cutoff equation can imply an arbitrarily small $s(\theta_t, Z)$, causing the cutoff rule to be too low. The cutoff that maximizes the bilateral surplus between the worker and the firm is generally not the same as

Separation and net present income. All-else equal, the model predicts that a retained worker's net-present income is strictly higher than that of a separating worker. Let $V^R(\theta_t, z_{1,t}, z_{2,t})$ and $V^N(\theta_t, z_{1,t}, z_{2,t})$ denote these respective payoffs.

$$\underbrace{S_t(\theta_t, Z)}_{\text{Retained NPV earnings}} > \frac{1}{1-\delta}\mathbb{E}(s^\star(z)|z_{1,t} < z \leq Z) > \underbrace{V_t(\theta_t, z_{1,t}, Z)}_{\text{Separated NPV earnings}} \tag{1.8}$$
$$\text{where} Z = Z_t(\theta_t, z_{1,t}, z_{2,t}).$$

This will form the basis of an empirical test that I discuss in the identification section.

Reallocation and Signaling Content of Endogenous versus Exogenous Separations. The most interesting testable implication of the model pertains to the reallocation and signaling content following separations. If a separation is endogenous—i.e., the worker's ability was below the incumbent's cutoff—then the worker's ability should be lower than the ability of an otherwise similar retained worker. Moreover, according to Proposition 3, she should move to a lower type firm.

On the other hand, if the separation is exogenous—i.e., unrelated to the worker's ability—then the worker's ability should still be less, at least on average, then that of an otherwise similar retained worker. However, the model predicts that, if the number of remaining periods are sufficiently high, the worker will move to a higher-type firm. The reasoning is simple: as $T - t \to \infty$, assignment becomes an increasing and approximately Markov function of the posterior (z_1, z_2). Workers who are displaced during a lengthy employment spell have a strictly better posterior than when they started, so they move to a better firm.

The literature on asymmetric employer learning commonly predicts that displaced workers should suffer less adverse wage changes than laid off workers because the latter carry less stigma. However, the prediction that displaced workers move to *better* firms appears to be a unique consequence of the interaction between complementarity and asymmetric learning.

Turnover, age, tenure, and firm type. The model predicts that turnover is a function of age t, the uncertainty in ability measured by $\frac{z_{2,t}}{z_{1,t}}$, and incumbent firm type relative to ability measured by $\frac{\theta}{z_{1,t}}$. Ceteris paribus, turnover is increasing in firm type, increasing in age, and decreasing in ability uncertainty. The unconditional relationship between tenure and turnover is driven by these variables. I.e., workers with identical posteriors, firm types, and age are predicted to have the same amount of tenure.[11]

Wage dynamics. The wages of retained workers are always equal to the surplus of the marginally retained ability type. The marginally retained ability type increases with tenure, and thus so do wages.

Conditional on a retained worker's posterior ability, higher type firms pay lower wages, because the marginally retained worker is always below the full-information match, which guarantees by Assumption 2 that the marginally retained worker's surplus is decreasing in firm type. However, because higher-type firms have higher turnover rates, retained workers at higher type firms experience faster wage growth.

The wages of new hires, given in Equation 1.1, behave quite differently. A new hire's wage equals her expected first period of output, plus any additional expected rents during the subsequent employment spell. As a result, it is possible for new hire wages to be strictly larger than subsequent retention wages. Although this may initially seem counterintuitive, it is perfectly consistent with the observation that employers often pay large signing bonuses to their recruits. As the uncertainty in a worker's posterior declines ($z_{1,t}/z_{2,t}$ approaches 1), the difference between the average and marginal output in a firm should decline, and these signing bonuses should become less important.

Earnings dispersion. As is typically the case with symmetric learning models, the model predicts rising earnings dispersion with age. Each state branches into three possible successor states (separation, displacement, retention), necessarily increasing ex-post earnings inequality compared to the predecessor state.

[11]The equilibrium will produce matches in which workers with different posteriors match to the same firm. For example, in the infinite horizon case, a worker who is repeatedly retained by her first firm will have strictly higher cutoffs than a worker of the same age who has just joined the firm.

1.4 Identification

I will now describe how to identify the theoretical model under several scenarios with different degrees of observables and auxiliary assumptions. Suppose we observe firms indexed by $f = 1, ..., F$ and workers indexed by $i = 1, ..., N$. In what follows, I will write $f(i,t)$ to denote the firm of individual i in period t, θ_f to denote the type of firm f, $(z_{i,1,t}, z_{i,2,t})$ to denote the posterior ability of worker i in period t, and $Z_{i,f,t}$ to denote the cutoff that firm f uses on worker i in period t.

Assumption 3 (Empirical shocks)**.** A worker's empirical assignment is equal to the equilibrium assignment function plus an error $u^1_{i,t}$ such that

$$\ln \theta_{f(i,t)} = \ln \theta_t(z_{i,1,t}, z_{i,2,t}) + u^1_{i,t}.$$

$u^1_{i,t}$ is i.i.d. and independent of true ability z_i. Similarly,

$$\ln w_{i,f,t} = \ln w^R_t(\theta_t, z_{i,1,t}, z_{i,2,t}) + u_{i,t,2}.$$

$u_{i,t,2}$ is i.i.d. and independent of true ability z_i.

This relaxes the model to allow for idiosyncratic deviations from predicted equilibrium assignment patterns, which are then assumed to resume following equilibrium play.

The worker and firm exit rates are straightforwardly identified.[12] The model has two remaining unknowns: the period-surplus function $s(\theta, z)$, and the distribution over pre-market posterior ability. This section will present multiple complementary identification strategies, which all rely on (1) inferring firm-types, $\theta_{i,t}$, and (2) inferring workers' posterior abilities throughout time, $(z_{i,1,t}, z_{i,2,t}$.

Assume that we observe panel data on an infinite number of workers workers $i = 1, ..., \infty$ observed in periods $t = 1, ..., T$ of their labor market history. For each worker-period ob-

[12] Random exit is the only source of time-discounting. Adding pure time preference or secular productivity growth could be easily accommodated.

servation, we can observe the mutually exclusive 'mobility outcomes' where the worker is retained ($R_{i,t}$), separated ($S_{i,t}$), or displaced by a firm exit ($D_{i,t}$) at the end of the period.

1.4.1 Identifying Posteriors upon Entry

The prior over z is the standard uniform distribution. However, workers may have public information that is revealed about them upon entry, but before they begin matching with firms.

Assume that the econometrician observes all of the initial public information about workers when they first enter the market unattached. Then I will show that it is possible to identify the mapping between this information, and the pre-market posteriors, based on the relationship between net present income and initial information.

Assume that every piece of information is binary.[13] This could include knowledge of GPA thresholds or certifications, for example. Each individual piece of information need not directly imply a cutoff with-respect-to z—for example, graduating from Harvard could be a positive signal for women and a negative signal for men. My strategy will hinge on first identifying the mean ability attributable to each distinct set of information. Let $X = \{\mathbf{x}_k\}_{k=1}^{K}$ enumerate the set of positive-probability realizations of pre-market information.

My strategy requires first being able to observe average ability, or some other percentile of the ability distribution, for each realization of $\mathbf{x}$.

Proposition 5. Suppose that, for each $\mathbf{x}_k$, the econometrician either observes the qth quantile, q_k, of ability, or observes the probability $\rho_k \in (0, 1)$ of ability exceeding some fixed percentile in the unconditional ability distribution. Then the pre-market posteriors $\{z_1(\mathbf{x}_k), z_2(\mathbf{x}_k)\}_{k=1}^{K}$ are non-parametrically identified.

See proof on page 77.

[13] The assumption that workers enter the market with a (potentially non-standard) uniform ability means that this is without loss of generality

1.4.2 Identifying Post-Entry Posteriors and Cutoffs

Given the posterior upon entry identified above, we can identify the entire sequence of cutoff rules used and posterior ability during a worker's career using mobility outcomes and observable retention probabilities.

A worker's current posterior $(z_{i,1,t}, z_{i,2,t})$ is a function of her previous posterior, last period's mobility outcome (separation, retention, or neither if the worker was dispalced or began unattached), and last period's retention probability. The current retention probability $r_{i,t}$ is a function of the incumbent firm and the worker's current posterior. Therefore, we can identify each worker's posterior dynamically using the following procedure.

First, set $t = 1$ and let the initial posteriors identified above correspond to $(z_{i,1,t}, z_{i,2,t})$. We will seek to build a sequence of posteriors and cutoffs, with the latter defined by convention to equal the lower-bound of the posterior in periods where the worker is initially unattached.

For each i with an incumbent firm, identify

$$r_{i,t} = Pr(\text{separate}|\text{separated or retained} f(i,t), z_{i,1,t}, z_{i,2,t}).$$

By the uniform assumption, the cutoff rule is $Z_{i,t} = (z_{i,2,t} - r_{i,t}(z_{i,2,t} - z_{i,1,t}))$. If i has no incumbent firm, $Z_{i,t} = 0$. Then, iterate the posterior forward using

$$(z_{i,1,t+1}, z_{i,2,t+1}) = \begin{cases} (z_{i,1,t}, z_{i,2,t}) & \text{if } i \text{ neither separate nor retained at } t \\ (z_{i,1,t}, Z_{i,t}) & \text{if } i \text{ separates at } t \\ (Z_{i,t}, z_{i,2,t}) & \text{if } i \text{ is retained at } t. \end{cases}$$

If we continue iterating for all t, we will have fully identified the sequence of posteriors for each worker.

1.4.3 Identifying Firm Types

There are two strategies for identifying firm types, which all require some sort of scalar normalization, so I assume that the lowest firm type equals 1.

Ranking Firms by Performance. Suppose that the econometrician observes a firm's accounting profit $\pi_{f,t}$ defined as total output minus payroll costs. Assume that the measured profit equals the equilibrium profit plus a mean-0 i.i.d. error, and that firms are observed for infinitely many periods. Then a firm's expected equilibrium output and profit are both proportional to θ_f^{ϕ}. Hence, firm types satisfy the following regression equation:

$$\ln \pi_{f,t} = \alpha_0 + \phi \ln \theta_f + u_{i,t}^3, \qquad u_{i,t} \sim i.i.d., \mathbb{E}u_{i,t} = 0.$$

Assume without loss of generality that firms are ranked in ascending order of expected profit. Then $\alpha_0 = \mathbb{E}\left(\ln \pi_{k,t} | f = 0\right)$ and $\theta_k = \mathbb{E}\left(\ln \pi_{f,t} | f = k\right) - \mathbb{E}\left(\ln \pi_{f,t} | f = 0\right)$.

One advantage of this approach is that it will allow us to directly test Proposition 3 by estimating the equilibrium assignment function in a model-free way and comparing it to the model's predictions.

Ranking Firms by Recruitment. Workers' posterior abilities are identified. Firm types can therefore be identified based on the posteriors of the workers that they hire. From Proposition 2, we know that firm assignment is homogeneous of degree 1. Hence,

$$\theta_t(z_{1,t}, z_{2,t}) = z_{1,t}\theta_t(1, \frac{z_{i,1,t}}{z_{i,2,t}})$$

which implies

$$\ln \theta_{f(i,t)} = \ln z_{i,1,t} + \ln \theta_t(1, \frac{z_{i,1,t}}{z_{i,2,t}}) + u_{i,t}^1.$$

Taking expectations over the workers hired into f implies

$$\ln \theta_f = \mathbb{E}\left(\ln z_{i,1,t} + \ln \theta_t(1, \frac{z_{i,1,t}}{z_{i,2,t}}) | i \text{ hired into } f \text{ at } t\right).$$

Fixing the ratio $z_{i,1,t}/z_{i,2,t}$, each firm's type can be identified up to a constant of proportionality by varying the observed $z_{i,1,t}$ of new hires. Combined with the assumption that $\theta_0 = 1$, firm types are identified.

1.4.4 Identifying the Surplus Function

Consider the wage equation $\ln w_{i,t} = \ln s(\theta_{i,t}, Z_{i,t}) + u_{i,t,2}$. Dividing by $\theta_{i,t}$ and using homogeneity yields

$$\ln w_{i,t} = \phi \ln \theta_{i,t} + \ln s(1, \frac{Z_{i,t}}{\theta_{i,t}}) + u_{i,t,2}.$$

We now effectively observe $Z_{i,t}$ and $\theta_{i,t}$. Let $W(\theta, \zeta) = \mathbb{E}(\ln w_{i,t} | \theta_{i,t} = \theta, \frac{Z_{i,t}}{\theta_{i,t}} = \zeta)$.

Thus, we can identify ϕ as

$$\frac{\partial W}{\partial \theta},$$

and then we can identify the function

$$\ln s(1, \zeta_{i,t}) = W - \phi \ln \theta_{i,t}$$

over the support of $\zeta_{i,t} = Z_{i,t}/\theta_{i,t}$. Thus, the surplus function is non-parametrically identified for all observed values of $\zeta_{i,t}$.

1.5 Testing

This section derives a series of empirical tests of the model that can be implemented using the same kinds of data as discussed in Section 1.4.

Direction of mobility. The model's most distinctive predictions are that separating workers move to lower-ranked firms, while displaced workers mobility is *less* downward, and potentially upward (as discussed, a sufficient condition is if T large). This can be immediately tested by computing the average changes in firm type by type of separation.

Over-placement. The model predicts that new workers over-place—that is, place higher than the static expected surplus-maximizing match $\bar{\theta}_{i,t} = \arg\max_{\theta} \mathbb{E}\left(s(\theta, z) | z_{i,1,t} < z \leq z_{i,2,t}\right)$, but that the degree of over-placement is decreasing in age and the ratio $z_{i,1,t}/z_{i,2,t}$.

The surplus function and workers' posteriors are identified, so $\{\theta^\star(z_{i,1,t}, z_{i,2,t})\}_{i,t}$ is identified. Thus, the over-placement prediction can be immediately tested by comparing $\bar{\theta}_{i,t}$ to $\theta_{i,t}$ across worker-year observations.

Exogenous exit. The model assumes that exits of firms and workers are exogenous to their types. There are several forms of endogenous exit, and failing to model them will generally result in incorrect model predictions, and in some cases require adjusting the identification strategy.

Under the null hypothesis that the model is correctly specified, testing exogenous firm exit is equivalent to testing the statistical independence of firm exit from identified firm type. Testing exogenous worker exit is more complicated. First, we can test the statistical independence of the identified posterior ability from worker exit. However, even if exit is independent of posterior ability, it could nonetheless be dependent with true ability. Thus, another test of the model that speaks directly to this issue is to test whether the unconditional distribution of talent, according to the distribution of the posteriors, remains uniform across each value of age t.

Turnover profiles. The model predicts that turnover rates are an increasing function of three arguments: t, $\theta_{i,t}/z_{i,1,t}$, and $z_{i,2,t}/z_{i,1,t}$. All three arguments are identified, and retention is observable, so the predicted relationships are directly testable.

1.6 Conclusion

This paper developed a new theory of matching and retention in a labor market with both firm-worker complementarity, and asymmetric employer learning leading to adverse selection. The majority of asymmetric employer learning papers following the tradition of Greenwald (1986) abstract from firm-worker complementarity. Accordingly, improved transparency of ability simply transfers rents and increases inequality without raising ex-ante surplus. The main contribution of my model, by adding firm-worker complementarity, is to give labor economists a better framework for assessing the potential for transparency-enhancing institutions to raise economic surplus.

CHAPTER 2

Adverse Selection and Reallocation Patterns in Law

2.1 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?

To help answer these questions, this paper documents several interesting features of assignment dynamics based on a historical dataset covering US lawyers form 1930-1963. I developed the data 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. The stakes of efficient assignment are particularly high for lawyers because of their large investments in human capital, and the wide heterogeneity in the type of work that law firms perform. Moreover, law firms tends have salient rankings, where top firms perform more sophisticated and lucrative work and tend to pay out-sized salaries in order to attract the most talented workers. The reallocation of lawyers through this intense heirarchy is an interesting phenomenon in its own right.

Using my data, I show that lawyers begin their career by assortatively matching into firms based on the quality of their schools. I classify separations after the exit of a firm as exogenous, and separations after survival of a firm as endogenous, and I find that endogenously separating lawyers are significantly adversely selected on latent measures of ability, consistent with asymmetric employer learning, and move to smaller firms where their colleagues are of lower average ability. Meanwhile, displaced lawyers show almost no evidence of selection, but nonetheless move to larger firms where their colleagues are of higher average ability. These findings support the model of Chapter 1, where firm-worker complementarity and asymmetric employer learning interact to generate assortative initial assignments, adverse selection, and lopsided reallocation dynamics.

My results seem to contrast with a commonly found tendency for inter-firm mobility to flow up the 'job ladder'. In fact, the lens provided in Chapter 1suggests that separating lawyers are adversely selected, and generate more economic surplus when matched with lower-ranking firms because of firm-worker complementarities in production. Conversely, retained workers are positively selected, so moving up in surplus requires moving up in firm rank—but not until an exogenous shock to their firm frees them from the adverse selection problem.

Having summarized the main results, I will end the introduction with a review of the related literature. The rest of the paper will be divided as follows. Section 2.2 describes the data, and Section 2.3 presents the empirical results. The final section concludes.

2.1.1 Related literature

Asymmetric employer learning. 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.[1] 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

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

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.

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

There is fairly abundant empirical evidence for this view of the job ladder based on comparing the transitions of poached workers to workers who are displaced into unemployment before reemploying. 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).[2]

The results of this paper do not support the existence of a unanimous job ladder in law. Instead, they support the idea, as in Lise and Robin (2017) and Eeckhout and Kircher (2011), that firm-worker complementarity results in workers of different abilities having a different job ladder or surplus-ranking of firms. If separating lawyers are adversely selected,

[2]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.

and lower-ranking firms have a comparative advantage in producing with low-ability lawyers, than movements down in firm-rank may constitute movements up the job ladder. Indeed, Haltiwanger et al. (2018) find that the evidence of a unanimous job-ladder is relatively weak among college-educated workers, which to some degree anticipates the results on lawyers in this paper.

2.2 Data and Background

My main data consist of linked entries in the annual *Martindale-Hubbell* professional directories covering US lawyers for all years between 1931 and 1963. I also match these data to deanonymized 1940 Census microdata, which I mainly use to infer permanent income from housing expenditure. *Martindale-Hubbell* (hereafter MH) is an information services company whose predecessor firms, *Martindale's* and *Hubbell's*, were founded in the mid-1800s and then merged in 1931. MH's principal products are biographical information on lawyers and legal digests.[3] Data from the MH directories have been used in several previous studies in economics and empirical legal studies.[4] 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.[5]

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

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

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

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

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 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.[6] 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 3.1.

1940 de-anonymized Census microdata. I match the MH data to the 100% Complete Count 1940 Census data from IPUMS. My main use of the Census data is to measure housing expenditure as a proxy for permanent income. To perform the matching, I extracted all the individuals from the Census whose listed occupation indicated a high likelihood of being a lawyer, and then used fuzzy matching on name, location, and age to match them to the MH data. If the individuals I failed to match are "unmatched at random," then dropping

[6] 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.

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.

My measure of housing expenditure contains 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.[7]

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.[8] Unlike in modern law firms, which typically feature four positions—associates, non-equity partners, equity partners, and permanent counsel—most

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

[8]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.

group practice lawyers in the sample period were identified simply as "members," or "partners."[9] Compensation agreements involved fixed profit-shares. Because there is not any public data on these agreements, no one knows exactly how unequal or complex they tended to be. Most biographies of law firms indicate that profit-shares were scaled with seniority or experience, and that occasionally one or two of the most senior partners would collect twice as much as everyone else (see Swaine (1948) and Bronson (1980)). In addition to these shares, partners were sometimes promised a minimum guaranteed base salary, as was the case when Cravath attempted to hire the famous John W. Davis (Harbaugh, 1973).

Summary statistics on the main variables used in the analysis are included in Table 3.1. 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.[10] 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, separation after firm exit, separation after firm survival, 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.[11] 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

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

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

[11]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.

lawyer i belongs to colleague set $\mathbf{c}_{i,t}$ in year t and the $\mathbf{c}_{i,t+1}$ in year $t+1$.[12] 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.[13] 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.[14]

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

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

[13] 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.

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

below 50%.[15]

Separations are distinguished by whether they coincided with the exit of the lawyer's firm, with exits being more likely to reflect an exogenous shock to the firm's business. Using these measures, I will now rank firms and study the dynamics of firm rank under the different types of mobility.

2.3 Empirical Evidence

I will now use the data to establish several facts about how lawyers match with firms and reallocate over time, with the intention of testing the model of Chapter 1. I will show that lawyers assortatively match into firms based on the quality of their law school, that separating lawyers move to lower-ranking firms if their incumbent firm survives, and higher-ranking firms if their incumbent firm exits. Finally, I will show that the first type of workers are significantly adversely selected, while the second type are not selected. To establish each fact, I will rely heavily on ranking schools and firms.

2.3.1 Scoring schools and ranking firms

I create a detailed measure of law school quality, LSQ. Then, using the principal in the model of Chapter 1that higher-type firms endogenously recruit workers with better signals of ability, I will rank firms on the basis of their expected new hires' LSQ. Thus, endogenous sorting patterns reveal primitive differences between firms. Because the firms in my data are small, I will predict new hires' LSQ based on firm characteristics rather than directly based on the firm's identity.

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 discrimina-

[15]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.

tion.[16] Moreover, I am only concerned with the signaling content of law school pedigree, inclusive but not exclusive to causal effects. 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.[17]

2.3.2 Measuring Law School Quality

I construct my own continuous measure of school quality, LSQ, based on a comparison of how each school's alumni fared in two outcomes during the sample period: 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 two fixed effects, after normalizing each of them into a Z-score.[18] To corroborate the LSQ measure, I will 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. 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.

I do not have law school data for lawyers who exited the sample prior to 1939—about

[16] 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).

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

[18] I experimented with using factor analysis to choose suitable weights for the effects and found them to be very close to a simple average.

17% of lawyers in the sample. For everyone else, an omitted law school implies 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. Readers who are not interested in the details of the LSQ estimates can skip the details below.

Details on LSQ estimates. 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.[19]

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, market size, and temporal trends. For net worth, I need to adjust for secular increases in 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 3.1 for an example and note that the intervals expand with inflation). I take the midpoint of the

[19] 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.

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.[20] 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.[21] Figure 3.2 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.[22]

Net worth ratings do a very good job of separating the lower half of schools. However, net worths are top-coded 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 dividing each measure by its standard deviation taking a simple average.

[20] 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.

[21] 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.

[22] 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)

2.3.3 Ranking firms.

With the LSQ measure in hand, the second step of the procedure forms an index of colleagues' characteristics based on the LSQ of the new lawyers their firm is predicted to hire. 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, $r_{i,t}$, is then transformed into a ranking among all other lawyers working at firms in the same year:

$$r_{i,t} = \frac{1}{N_f} \sum_{j=1}^{N_t} \mathbf{1} \left(\widehat{LSQ}_{j,f,t} < \widehat{LSQ}_{i,f,t} \right) \tag{2.1}$$

To validate this method, I show that estimated firm ranks are powerful predictors of career success. I consider three outcomes: log housing expenditures, 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 3.2. This does not necessarily imply that firms cause success. However, it is consistent with my preferred interpretation that complementarity drives assortative matching, and thus reflects a component of lawyers' abilities that are not reflected in LSQ.

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 necessarily that firm rank *causes* wealth, but rather that equilibrium assortative matching between firm and worker types causes firm rank to reflect some of the residual ability not explained by LSQ.

2.3.4 Empirical Facts on Matching and Reallocation

Fact 1: Assortative Matching by Quality of Law School. I will now show that lawyers assortatively 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 3.3, 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. One of the more immediate reasons why a high-type firm will be hesitant to hire a low-type worker is overhead costs. I have reviewed historical data on the overhead expenses of professional law firms during this time period and found that, indeed, overhead rises on a per-attorney basis with the stature of the firm. Figure 3.3 contains a reprinted table of per-lawyer overhead expenses compared to per-lawyer income across different net income groups for a sample of New York County lawyers surveyed in 1933 and reported in the American Bar Association's Economic Survey of the Legal Profession. The reason that overhead tended to increase with income is that higher-earning firms tended to recruit more capital and labor (typewriters, libraries, secretaries, stenographers, etc.) to support each of their individual lawyers.

Fact 2: Changes in firm rank after separation. In Table 3.4, I have regressed changes in firm rank on mutually exclusive indicators for separation following firm survival, separation following firm exit, and retention (the omitted category). The first column shows that lawyers who separate when their firm survives lose an average of 6 percentage points in rank, while lawyers who separate after their firm exits *gain* an average of 3.7 percentage points. However, separation rates can differ by firm rank, and a potential concern would be that if high-ranking firms had higher separation rates (they do) then mean reversion could create mechanical downward mobility. 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 separation 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 the two types of separations.

The evidence suggests quite robustly that lawyers lose rank when they separate from surviving firms and gain rank when they separate from exiting firms. These findings are opposite to what we would expect from the standard job ladder literature's predictions about poached versus displaced workers.

Fact 3: Adverse selection. Why do separating lawyers move to lower-ranked firms if their previous firm survives, and better-ranked firms if their previous firm exits? I will now present evidence that the former types of separations involve adverse selection. On the other hand, and as one might expect if firms exit because of exogenous shocks, the other set of lawyers appear to be neither negatively nor positively selected.

To test for selection on unobserved ability, I will again use the a ratings published by MH. Because lawyers do not qualify to receive an a rating until they have 10 or more years of experience, future a rating attainment can be thought of 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. Thus, the information about future a ratings attainment contained in a separation outcome reflects whether separations select on ability.

This type of latent-variable approach is a canonical method for testing whether firms privately learn about their employees' talents. Several papers starting with Gibbons and Katz (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).[23] 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 separates following firm exit, separates following firm survival, or is retained, as well as other relevant aspects of her job history. 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 3.5 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 separating lawyers are adversely selected when their original firm survived, but not when their original firm exits, strongly suggests that the former type of separation reflects a choice by the incumbent firm to selectively retain high-ability workers. It is not surprising that adversely selected lawyers are observed moving down in firm rank. However, it

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

is quite counterintuitive to see that lawyers whose firms exit tend to move to higher-ranking firms, despite no evidence that are positively selected. This is precisely the phenomenon explained by Chapter 1, in which workers become increasingly revealed ex-post to be *under-placed* at their incumbent firm, but cannot receive attractive offers to move up the ranks because of the adverse selection problem. Once an exogenous shock arrives to terminate the firm, they become free to move up the ranks.

2.4 Conclusion

This paper presented novel historical evidence on the matching and reallocation patterns in law. The assortative matching of lawyers by law-school quality supports a theory of productive complementarities between workers and firm technologies. The adverse selection of separating workers whose previous firms survived supports a theory of asymmetric employer learning about ability. I showed that lawyers reallocate *down* in firm rank when their previous firm survives, but *up* in firm rank when their previous firm exits the market, despite no evidence that the latter group is selected. The strength of the model in Chapter 1is that it requires no added forces to explain these last two facts. The interaction of firm-worker complementarity with asymmetric learning results in lopsided mobility, where endogenous separations lead down the firm heirarchy, while exogenous separations lead up.

CHAPTER 3

Retention and Optimal Paysetting in the Federal Government

3.1 Introduction

Whereas perfectly competitive labor markets admit a single *going wage*, employers in the real world have discretion to choose different levels of pay in exchange for different hazards of quitting. How can employers identify the optimal tradeoff when choosing a compensation policy?

In this paper we build a framework for studying the long-term consequences of different compensation policies on payroll, workforce composition, and retention. We use our framework to study a massive federal wage reform affecting over a million US workers and cumulatively costing over a hundred billion dollars: the Federal Employees Pay Comparability Act, or FEPCA. We use the government's personnel data combined with a natural experiment inherent to FEPCA in order to estimate a dynamic retention model that allows us to assess the impact of additional compensation, such as that provided by FEPCA, on retention.

Some labor markets are dynamic enough that employers may be expected to acquire optimal compensation policies via spontaneous trial and error. But many labor markets are too bureaucratic (or concentrated) for trial and error alone to probe the relevant policy space. Our structural framework is therefore particularly helpful for studying the decisions of bureaucratic employers.

Our paper both exploits FEPCA as an identification strategy and provides the first estimates of its effectiveness. After decades of operating on a national pay-schedule that ignored spatial wage differences, the federal government passed FEPCA in order to provide *locality pay* to workers as a function of benchmark pay indices. These indices were computed at fine-grain occupation and spatial levels. Our identification strategy exploits an idiosyncratic feature of the locality pay formula where these fine-grain indices were spatially averaged in order to produce a *single* extra pay rate for each locality.[1]

This spatial averaging created variation in locality pay based on the plausibly exogenous composition of the local federal workforce. Certain technical occupations, like engineers, were found to be underpaid all over the country. Hence, a worker's locality pay was largely determined by the fraction of the local federal workforce that belonged to one of these occupations. Underpaid engineers in Indiannapolis received low locality pay, while well-paid secretaries in Houston (home to NASA) received high locality pay.

Formally, our approach is to specify a dynamic discrete choice model of quitting the federal government which maps structural parameters capturing the substitutability of different employment opportunities into both observed and counterfactual quitting probabilities. We estimate the model using observed quitting rates in localities differentially affected by FEPCA. The model can then be used to simulate the effects of alternative pay policies on retention, payroll, and workforce composition.

Prior literature. Our quitting model is a variant of the Dynamic Retention Model (DRM) originally pioneered by Gotz and McCall (1983) (henceforth, GM). GM formulates quitting as the solution to a once-and-for-all optimal stopping problem based on an ongoing comparison of stochastic net present income streams of outside work contained to the option value of being retained for one additional period. GM used their model to predict how recently contemplated reductions in retirement benefits would affect the reenlistment decisions of Air Force officers. They find that a 5% unanticipated increase in compensation in 1976 and 1977 would have raised two-year retention rates from 83% to 86% for ROTC pilots and from 74% to 79% for non-pilot officers. Daula and Moffitt (1995) (DM) uses a very similar framework to study the reenlistment decisions of army infantrymen finishing their second and third

[1]Initially, there were 32 federal localities. Over time, the government added additional localities and changed the boundaries of the existing ones. As of 2011, the last year of our data, there were 35.

terms between 1974 and 1984, finding that a 10% increase in army pay would have increased the first-term reenlistment rate from 33% to 40%, and the second-term reenlistment rate from 63% to 71%.[2] Knapp et al. (2016) use a DRM to study compensation and retirement eligibility of Chicago public school teachers. Variants of the DRM model continue to be used for workforce management in the public sector (see Asch et al. (2013) for a recent overview of RAND's DRM model).

At its inception, the DRM competed with several other modeling approaches for helping government agencies predict how changes to their retirement policies might impact their ability to retain employees. The main alternative was the Annualized Cost of Leaving (ACOL) model (e.g., Warner and Goldberg (1984)). The ACOL setup was similar to the DRM setup except workers were assumed to compute the value of being retained for one additional period by assuming, in a time-inconsistent manner, that they would never choose to quit in the future. Ignoring the option value of being able to quit in the future allowed the ACOL model to be computed more rapidly at the expense of making less realistic forecasts—a tradeoff that has become less desirable over time.

Given that corporations typically have formal compensation policies (see Baker et al. (1994) for a famous example), it is somewhat surprising that the literature appears not to feature a single application of DRMs in the private sector. This is not for a lack of interest. A large number of published articles in the operations research literature study optimal compensation policy in a Markov setting that closely matches our model, but these articles do not actually estimate the parameters governing workers' behavior. Hence, data availability seems to be the likeliest explanation for why all of the DRM applications have been in the public sector.[3]

The literature on pay and retention also features static reduced form models that predict quitting as a function of contemporaneous measures of inside versus outside pay.[4] Static

[2]Reenlistment decisions occured every four years.

[3]See Rao (1990) for one of the earlier examples in operations research.

[4]See Carrell (2007) who finds that inflexible Air Force pay systems lead to higher quit rates in locations with higher wage differentials, and Borjas (1982) who concludes that federal employees could be paid significantly less at minimal cost to retention.

retention models such as the above two papers do no usually attempt to isolate the direct effect of contemporaneous pay on retention from its correlation with future pay. Hence, reduced form pay elasticities don't identify a policy-relevant parameter when pay is serialy correlated. DRMs have the advantage of fully-specifying how the timing and uncertainty of current and future compensation change the propensity to quit.

The downside of most DRMs in the literature following GM and DM is that they tend not to address potential endogeneity between pay and unobserved outside options. Our DRM retains the seminal features of the GM and DM papers, but uses what we consider to be a more plausible identification strategy that exploits arbitrary unevenness in the effects that FEPCA had on federal workers' pay.

Both GM and DM allow for the presence of observable characteristics that simultaneously change the value of inside and outside work. Their identification hinges on a subset of instrumental variables that that are, implicitly or explicitly, excluded from the value of outside work. In GM's application, pay and promotion probabilities were a legally mandated function of experience and current pay grade, which are both used as instruments. In DM's application, the instruments include cohort-level changes to the retirement system and pay schedule, and the individual's time in military service. Both the idiosyncratic variation used by GM, and the aggregate policy-induced variation of DM, are potentially endogenous. Compared to these two papers, our paper clarifies the distinction between control variables and instruments for pay. We allow for the meritocratic or *basic* component of internal pay and an index of external pay to have arbitrary correlation with external wages. We assume that, conditional on these two components, the remaining component of pay that is induced by the locality pay policy is exogenous.

The second contribution of our framework is to model endogenous changes to workforce composition under counterfactual policies.Whereas the past literature has examined counterfactual retention rates of a static cohort of workers, we study the payroll, retention, and composition of a workforce that evolves endogenously via the replacement of departing workers with simulated new hires. This helps us account for a tendency of policies to change the composition of the workforce on both observables and unobservables.

This paper is the first to quantitatively estimate the effects that FEPCA has had on the government workforce. In anticipation of FEPCA, Lewis and Durst (1995) use a one percent sample of data on 1985-1989 federal employees to test whether areas with larger pay differentials experienced higher turnover rates. They found no evidence confirming higher quit rates in localities with higher wage differentials, albeit with limited statistical power. In comparison, we find a large and statistically significant partial correlation between quit rates and occupation-by-location outside-wage indices.

Many of the papers studying retention in the public sector have sought to better understand the role of non-pecuniary factors. Borjas (1982) considers how the political influence of different agencies affected the quit rates of their employees. Bolton et al. (2018) show evidence that presidential elections result in excess departures of employees from agencies with views opposed to the winner. Our results suggest large individual, time-invariant heterogeneity in workers' evaluation of the non-pecuniary benefits of federal service compared to their outside options.

Our estimates provide crucial insight into a heated policy debate over how to improve federal civil service compensation policies. The Heritage Foundation recently published a report claiming that federal workers are overpaid by 16% relative to similar private sector workers, and proposed to cut this premium by shrinking retirement benefits and automatic pay increases.(Greszler and Sherk, 2016) Others have cited the relatively low separation rates of federal employees as evidence that they are paid too much. However, Utgoff (1983) explains that neither parity in pay nor quit rates is the relevant policy objective.

> Firms choose compensation levels that minimize compensation and turnover costs, and this minimum depends, among other things, on the size of the firm and the fixed costs of hiring and training. To be efficient, the government must also select an optimal quit and compensation combination. Equal or comparable compensation is not an efficient principle.(Utgoff, 1983, p.396)

Our paper helps identify policies that obtain a cost-efficient combination of quits and compensation. The rest of the paper is organized as follows. Section 3.2 describes the over-arching Markov framework for modeling and assessing compensation policy, describes the problem of pay being endogenous to unobserved outside options, and lays out our control function approach to solving this problem. Section 3.3 describes the background and data. Section 3.4 contains additional parametric modeling restrictions and describes how we estimate the model. Section 3.5 concludes.

3.2 Theoretical Framework for Optimal Compensation

3.2.1 Managing a Workforce under Exogenous Staffing Demands

In our framework, the demand for staff is exogenous, and the planner's objective is to meet demand at minimum cost. The employer has to continually fill a set of exogenous *jobs*, $i = 1, ..., N$, with employees. When an employee quits, the employer must immediately pay a fixed cost ψ to find a replacement. The employer commits in advance to a dynamic compensation policy, which influences the quitting decisions of the worker, who is forward-looking and subject to random preference shocks. The behavioral response of quitting to compensation policy is assumed (for now) to be known by the planner.

Markov replacement dynamics. Each employee-slot pair is described by an evolving state vector, $x_{i,t}$, reflecting both the characteristics of the employee, such as age or experience, as well as the characteristics of the job itself, such as conditions of the local labor market. The employer chooses and commits to a wage-setting policy d that determines compensation as a function of $x_{i,t}$, $w_{i,t} = d(x_{i,t})$. Employees choose whether or not to quit based on the wage policy and the evolving state variables, $p^q(x_{i,t}, d)$. When an employee quits, she is immediately replaced by a new employee with new characteristics $x_{i,t}$ drawn at random and potentially dependent on $x_{i,t-1}$.

In some ways, the framework resembles Harold Zurcher's optimal bus replacement problem in Rust (1987). However, rather than directly choosing replacements, the decision-maker chooses a costly policy that *induces* the rate at which employees are replaced. Unlike Rust's framework, which assumes that existing policies are maximizing a known objective and inverts them to identify state-specific costs, we are interested in appraising existing policies that are known ex-ante to be suboptimal. The replacement costs are an input to our framework.

Costs of a slot. The employer discounts future costs at rate δ. Let $p^q(x_{i,t}; d)$ denote the quitting probability of a worker conditional on the state and the policy. The expected costs at time 0 to the employer of maintaining a single slot i is, in recursive form,

$$c_{i,t} = c(x_{i,t}; d) = w(x_{i,t}; d) + \delta \left[p^q(x_{i,t}; d)\psi c^0(d) + (1-\delta)\left(1 - p^q(x_{i,t}; d)\right) \mathbb{E}\left(c(x_{i,t+1})|x_{i,t}\right)\right],$$

where $c^0(d) = \int c(x; d) dF_0(x)$ is the expectation of the net present cost of a slot with a new, randomly drawn worker. The employer's optimal policy minimizes average net-present costs per slot, $d_t^\star = \arg\min_{d \in \mathcal{D}} \sum_{i=1}^{N} c(x_{i,t}; d)$.

In practice, it is not trivial for an employer to discover the the causal relationship between compensation policies and quitting rates, $p^q(x_{i,t}, d)$. Conducting randomized experiments that span the set of all possible compensation policies is not feasible. Our approach is to impose additional theoretical structure (via a dynamic retention model) so that $p^q(x_{i,t}; d)$ can be identified without variation in d.

3.2.2 Dynamic Retention Model

In order to model $p^q(x_{i,t}; d)$, we now introduce a dynamic retention model (DRM) that models quitting decisions as an optimal stopping problem. Workers experience period payoffs equaling the sum of earnings and earnings-equivalent non-pecuniary benefits, and seek to maximize net present payoffs.[5] A worker i in year t has observable characteristics $z_{i,t}$, such as age, experience, education, and compensation history, time-varying unobservable characteristic $u_{i,t}$, and a time-invariant characteristic v_i. She is subject to a known policy d that pays her a wage, $w_{i,t}$, while employed internally:

$$w_{i,t} = d(z_{i,t}, u_{i,t}).$$

[5]Similar to Keane and Wolpin (1997), this specification is consistent with a more primitive environment in which workers' have access to complete markets and have preferences that are separable between labor allocation and consumption.

After receiving her wage, the worker draws exogenous i.i.d. shocks to the value of staying $\epsilon^1_{i,t}$ and leaving $\epsilon^2_{i,t}$. After observing the shocks, she decides to stay for an additional period, or quit. If she quits, she receives terminal outside option

$$V^o(z_{i,t}, v_i, u_{i,t}) + \epsilon^2_{i,t}.$$[6]

If she stays, she receives $\epsilon^1_{i,t}$, and then enters next period employed with a new $z_{i,t+1}, u_{i,t+1}$ according to Markov transition kernel $P(z', u'|z, u)$. Her quitting decision is therefore an optimal stopping problem satisfying the Bellman equation

$$\begin{aligned} V(z_{i,t}, v_i, u_{i,t}, \epsilon^1_{i,t}, \epsilon^2_{i,t}) = d(z_{i,t}, u_{i,t}) + \\ \delta \mathbb{E}_{\epsilon^1,\epsilon^2} \max \left\{ V^o(z_{i,t}, v_i, u_{i,t}, \epsilon^1, \epsilon^2), \mathbb{E}\left[V(z_{i,t+1}, v_i, v_i u_{i,t+1}, \epsilon^1, \epsilon^2) | z_{i,t}, u_{i,t} \right] \right\}. \end{aligned} \tag{3.1}$$

One of the state variables in $z_{i,t}$ is the worker's labor market experience. We assume that there is some terminal experience level at which the worker is guaranteed to quit. This allows us to solve the Bellman equation by applying backward induction from the terminal experience level.

Shocks. Each worker's v_i represents a time-invariant individual affinity for outside work, relative to inside work, and is drawn i.i.d. from a known and fixed distribution at the time the worker enters the workforce. The i.i.d. shocks $\epsilon^1_{i,t}$ and $\epsilon^2_{i,t}$ are assumed to each follow Gumbel distributions with equal variances. Under this assumption, their difference $\epsilon_{i,t}$ follows a mean-0 logistic distribution with unknown standard deviation σ_ϵ. The standard deviation of the Gumbel shocks is the model's numeraire.

Value function and quit probabilities. The Gumbel assumption gives us a tractable analytical formula for the ex-ante value function, as in Rust (1987).[7] For convenience, we divide out the standard deviation of the shocks and work with $\beta = \frac{1}{\sigma_\epsilon}$. Let $\bar{V}(z, v, u) \equiv$

[6] We put more structure on the outside option in Section 3.4.

[7] Daula and Moffitt (1995) similarly include i.i.d. additive shocks, but assume that they are normally distributed, and thus must approximate the integral over the shocks' distribution.

$\mathbb{E}_\epsilon V(z, v, u, \epsilon)$.

$$\bar{V}(z, v, u) = \beta d(z, u) + \gamma + \delta \log \left(e^{\mathbb{E}\left[\bar{V}(z', v, u')|z, u\right]} + e^{V^o(z, v, u)} \right), \tag{3.2}$$

where γ is Euler's constant. Quit probabilities take a convenient log-odds ratio form:

$$\ln \frac{p^q(z, v, u; d)}{1 - p^q(z, v, u; d)} = \delta \left(V^o(z, u) - \mathbb{E}\left[\bar{V}(z', u')|z, u\right]\right). \tag{3.3}$$

This implies that a one-unit, ceteris paribus increase in the net present value associated with staying next period, $\mathbb{E}\bar{V}'$, will *reduce* the log-odds of quitting by β units. The following lemma helps us relate to past papers that often estimate the elasticity of quitting with respect to pay.

Lemma 7 (Compensation elasticity with Gumbel shocks)**.** Define the *total compensation elasticity* η as the elasticity of the quitting probability p^q with respect to the future ex-ante value function, $\mathbb{E}\bar{V}'$.

$$\eta = \beta(1 - p^q)\mathbb{E}\bar{V}'. \tag{3.4}$$

Proof. $\frac{\partial \ln p^q}{\partial \ln \mathbb{E}\bar{V}'} = \beta \frac{d \ln p^q}{d \ln \frac{p^q}{1-p^q}} \mathbb{E}\bar{V}' = \beta \frac{dp^q}{d \frac{p^q}{1-p^q}} \frac{\mathbb{E}\bar{V}'}{1-p^q} = \beta(1 - p^q)\mathbb{E}\bar{V}'$. $\square$

The control function assumptions. So far, we have imposed no restrictions on the process governing $u_{i,t}$ or how it enters $V(\cdot)$ and $V^o(\cdot)$, so the exogenous i.i.d. shocks introduced above are thus far redundant. We will now make a set of control function assumptions pertaining to $u_{i,t}$ and explain how it identifies the key causal parameter, β.

Assumption 4 (Separability)**.** Under the existing compensation policy d, wages $w_{i,t}$ given $z_{i,t}$ are invertible in $u_{i,t}$.

Under this assumption, $u_{i,t}$ can effectively be observed.

Assumption 5 (Exclusion restriction)**.** There exists a subset $z^1_{i,t}$ of variables in $z_{i,t}$ that do not influence the outside option, V^o. I.e., $\frac{\partial V^o}{\partial z_{i,t}} = 0$ with probability one.

Assumption 6 (Relevance)**.** $z^1_{i,t}$ influences . I.e., $\frac{\partial w}{\partial z^1_{i,t}} \neq 0$ with strictly positive probability.

Proposition 6. Suppose $v_i = 0 \forall i$. Then β is identified.

The proof comes directly from the pay elasticity formula in Equation 3.4. Conditioning on $u_{i,t}$ and the non-excluded variables, the instrument(s) can be varied to study the effects of raising total net present future compensation on quitting probabilities, and hence identify β. Of course, we are interested in the case where $v_i \neq 0$, and we are interested in identifying more than simply β, and to do this we will make stronger parameteric restrictions in Section 3.4. Nonetheless, these assumptions capture the core of our identification strategy.

Finding a credible instrument satisfying Assumption 5 and Assumption 6 is usually difficult. For example, Gotz and McCall (1983) assume that workers' pay grades affect their inside wages but not their outside options (conditional on other controls). But if some workers are in higher pay-grades than otherwise-similar peers, this might reflect unobserved differences in effort or skill. On the one hand, if a worker knows that she has relatively poor outside options, then she would naturally want to invest more in her career within the organization, which would suggest that V^o could be decreasing in pay-grade. On the other hand, if a worker is highly paid because of unobservably high skills, then she probably also has access to unobservably high outside options, suggesting that V^o should be increasing in pay grade.

Daula and Moffitt (1995) use aggregate changes to the national pay schedule and retirement benefits as their sources of exogenous variation. A possible drawback of using aggregate changes is that they could very well capture endogenous responses to unobserved aggregate shocks. An advantage of the FEPCA experiment is that much of the variation in locality pay appears to be truly arbitrary.

3.3 Background and Data

3.3.1 The General Schedule Pay System and the CPDF

We are specifically interested in compensation policy and retention in the General Schedule (GS) pay system, which manages Human Resources policies for over 1.5 million white-collar federal workers worldwide. To study this workforce, we obtained access to the GS system's administrative personnel dataset, known as the Central Personnel Data File (CPDF) from the Office of Personnel Management (OPM). Our sample contains all federal civil servants in the GS system from 1989 to 2011, excluding employees of the Department of Defense, United States Postal Service, or sensitive agencies and/or positions designated by the federal government. See Bolton et al. (2018) and Zhang and de Figueiredo (2018) for more details on the OPM dataset.

Workers in the GS system are ranked into grades 1-15 and steps 1-10. Their base salary or 'basic pay' is determined by an annual grade-by-step matrix. Higher grades reflect increasing skill and/or responsibility, and grade increases are promotions. Steps are, with rare exception, a deterministic function of time spent in a grade.[8] The GS workforce is divided into administrative geographic units known as *localities*, which are similar to metropolitan statistical areas, and fine-grain occupational categories based on an OPM classification system.

To perform our analysis, we had to combine two distinct modules of the CPDF: the status data, and the dynamic data. The status data report an annual snapshot of the workforce at the end of every fiscal year in September, and the dynamic data report individual personnel actions arriving in real time. The status data track basic pay, locality pay (from FEPCA), location, occupation, age, and experience. The dynamic data help identify quits.

Quits. For our analysis, we are interested in studying separations that are initiated by employees and are therefore informative of preferences for inside versus outside work based on the usual compensation policy. We therefore would like to abstract from involuntary separations such as layoffs and firings (because they don't reflect the worker's preferences). Fortunately, involuntary separations are very rare in the federal government. Actual firings almost never occur because of costly appeals proceedings. Agency restructurings often lead to 'reductions in force' which can sometimes result in separations, but more often result in internal transfers. The actual rate of involuntary separations from our sample is a mere 0.03%,

[8]Grade-increases are usually earned through competitive application for vacant jobs, but some workers are hired on so-called ladder jobs where promotion occurs automatically after a year or two unless performance is unsatisfactory.

compared to a typical range of 12% to 24% in the BLS Job Openings and Labor Turnover Survey (JOLTS) from the past decade. Our analysis drops anyone who was involuntarily separated at some point during their career.

Sometimes, employees volunteer to retire early in exchange for buyouts (known as voluntary separation incentives) that are offered first-come-first-serve by downsizing agencies. The choice to accept a buyout, conditional on its being made available, would be a very interesting source of variation, but unfortunately we do not have a good method to identify exactly to whom buyouts were available. We only observe the accepted offers. Therefore, we drop individuals who at some point accepted a VSI.

We drop a small number of individuals who at some point in their tenure have contradictory personnel actions that seem to be due to clerical errors. We also drop individuals who exit the status data without any indication of a quit. Our belief is that some of these individuals are transferring into the Department of Defense or United States Postal Service (in which case we cease to observe them unless and until they return to the scope of our data-set).

Age. To protect individual privacy, OPM injected error into the age variable in the status data. For each person-year observation, they first added a random integer drawn uniformly from $\{-2, -1, ..., 2\}$ to the raw age field, and then they binned it into a five-year interval. We observe the midpoint of the interval $(17, 22, 27, ..., 62, 67)$. Because the errors are i.i.d., individuals can have mutually inconsistent age bins over time. We opt to use only the first observed age bin for each individual and we explicitly account for the measurement error in the estimation.

Years of Service. A worker's years of service in the federal government is important for keeping track of accumulated pension benefits and therefore plays an integral role in our analysis. The CPDF contain data on the number of creditable years of federal service for the purpose of computing retirement benefits and leave accrual. Creditable years of service usually corresponds to the number of years the employee has been retained in the GS system. But it can occasionally reflect time served outside the GS system, such as in the military or national guard. Parts of our analysis will exclude people with extra creditable years of service.

3.3.2 FEPCA

The centerpiece of our identification strategy is exploiting FEPCA to get an instrument satisfying our control function assumptions in Section 3.2. FEPCA was passed in 1991 after stagnating federal wages became a recognized public policy issue.

The Volcker Commission. Paul A. Volcker chaired a National Commission on Public Service in 1989, in which policymakers met to discuss challenges facing the public sector workforce. The Commission concluded that wages in the GS pay system had failed to keep up with the private sector, and that government mission objectives were being compromised as a result. The Commission report writes,

> Unfortunately, after a decade of budget cuts and pay freezes, salaries of most federal employees are clearly lagging behind the private sector. According to the most recent survey conducted by the Bureau of Labor Statistics, the gap, on average, nationwide is 22 percent. Whether or not that figure accurately captures the differential for particular jobs and areas, there is no doubt that the pay gap has become a disincentive in both recruiting and retaining a high-performance work force. Many young Americans feel they can no longer afford to take a government job, while many civil servants can no longer ignore the call of private pay. (Volcker, 1989, p.37-38)

The Commission did not simply recommend increasing GS wages across the board. It advocated for a closely-tailored adjustment to wages that factored in both occupation and location.

> The pay rate for a government secretary working in Phoenix, for example, is much closer to that of the private sector than is the pay of a physician at the National Institutes of Health or an engineer at the Marshall Space Flight Center...

> These locality and occupational differences seriously undermine the fairness of the current civil service pay process, especially when all federal employees receive the same pay whether they work in low-cost areas such as Kansas City or Norfolk, Virginia, or high-cost areas such as San Francisco or Boston. (Volcker, 1989, p.38)

The government passed FEPCA in 1990 and mandated that the BLS would construct annual measures of pay disparities based on location and occupation, which would then be used in a formula to determine annual pay adjustments as a rate of basic pay. However, the formula did not allow workers in the same locality to receive different adjusted pay rates. Instead, each locality had a single annual locality pay rate based on the average pay disparity of the local workforce.

FEPCA became effective for most localities in 1994, but had emergency provisions to immediately increase wages in San Francisco and New York starting in 1991. Figure 3.5 (data taken from Monthly Labor Review, 2009) documents the slowdown in GS pay growth, followed by the implementation of FEPCA.

FEPCA provided detailed instructions for choosing locality pay rates. A set of 30 localities were chosen corresponding to the largest Metropolitan Statistical Areas by federal employment. All other areas were grouped into a single, expansive locality called the Rest of US .[9] First, the newly created Federal Salary Council (FSC) would estimate average wage gaps in each locality. This is the average percentage point difference between an outside measure from the BLS, and a worker's current federal wage. Second, the FSC would calculate an annual target rate in each locality. This was the percent increase in pay needed to reduce the average wage gap to five percentage points. That is, in locality l,

$$TARGET_l = (1 + WAGEGAP_l)\,/1.05 - 1$$

[9] Over time, areas are added to the list of localities if they have sufficiently high federal employment and if there is evidence that market wages deviate significantly from the Rest of US. The coverage of individual localities has also grown. Unfortunately, these changes appear to be to small to be used as a source of exogenous variation in wages. For our analysis, we use the most up-to-date locality pay area definitions as of 2017.

The locality pay rates were mandated to gradually increase as a fraction of these targets. Pay rates were supposed to start at 20% of target in 1994, and increase by 10 percent of target each year until reaching 100% in 2002. The FSC was required to recommend these rates to the President. But Presidents Clinton, Bush, Obama, and Trump have typically used their authority to increase rates by smaller factors than recommended by the FSC.

Figure 3.4 illustrates our own calculation of the locality pay rates based on the language of FEPCA and published wage disparities by the FSC, compared to actual loclaity pay rates, in eight selected localities. Because of the budgetary demands created by FEPCA, Congress began allocating less money to nationwide General Schedule pay increases. Figure 3.5 shows the trajectory of GS pay increases and locality pay increases over time.

It is this aspect in which the implementation has most importantly deviated from the original intentions of FEPCA. Since 1994, Presidents have allocated less than the mandated increase. As a result, pay disparities remain far above five percent in most localities. Nonetheless, the *relative* allocation of locality pay among localities, and the growth rates over time, have been in line with the original bill.

Figure 3.4 displays mandated versus actual levels of locality pay for four of the most and least affected localities. Houston, Los Angeles, New York, San Jose/San Francisco are in the first panel of Figure 3.4 while Columbus (Ohio), Kansas City, Pittsburgh, and the Rest of US (abbreviated as HO, NY, SJ, and ZX, respectively) are shown in the second panel (with a different scale). The mandated rates are calculated using published estimates of pay disparities.[10] The panels show that locality pay rates have grown steadily over the years, but tend to be about a third as large as the policy's prescription. The percentage shortfall between mandated and actual rates is very similar across localities.

FEPCA did not legislate its own funding, so Congress had to account for locality pay increases in the annual appropriations process. The most affected agencies were given extra funding, and when this was insufficient, they cut payroll costs by offering VSIs (the aforementioned buyouts) to older employees.[11]

[10] See the Pay Agent Reports

[11] We know this from analyzing the data ourselves and from discussions with federal employees.

Ultimately, FEPCA directly increased the pay of over one million employees at a cost on the order of a hundred billion dollars (2016 CPI) between 1994 and 2011. As of 2017, the annual cost is reported to be $25 billion (2018 Pay Agent Report).[12] It is difficult to imagine a wage reform of similar scale and scope.

3.3.3 BLS National Compensation Surveys

To help in our analysis, we obtained the National Compensation Survey (NCS) data that used to help choose locality pay rates. The surveys were designed specifically for use in FEPCA to measure wages for jobs in similar locations, occupations, and ranks as individual federal workers. The NCS data were collected from 1997 to 2010.[13] The BLS's methodology was to send surveyors to sets of randomly selected firms in geographical areas to collect data on the wages, hours, occupational category, and description of work of different local employees. Each surveyed worker was assigned the GS grade level commensurate with the scope of their work and their level of responsibility. All individual workers within the selected firms were interviewed. About half of the survey areas were large cities corresponding to individual localities, and the other half were small cities and urban communities comprising the Rest of US locality.

The purpose of using this data is to construct a variable capturing occupation-by-location measures of outside pay. Our identification strategy does not assume that locality pay was completley exogenous. Instead, it assumes that *conditional* on the outside pay measures (that were spatially averaged to determine FEPCA), the level and growth rate of locality pay is exogenous to workers' outside options. Our identification strategy assumes that, conditional on these fine-grain measures, residual variation in locality pay rates are driven by differences in workforce composition that are exogenous to workers' unobserved quitting propensities.

[12] The costs are much higher if one considers non-GS employees for whom the government has volunteered to match locality pay rates.

[13] The NCS data changed classification systems in 2006. The 1997-2006 data can be downloaded from and the 2006-2010 data can be downloaded from .

Constructing our outside wage index requires us to map the NCS survey data to individual workers in the CPDF. We first aggregate the NCS survey data by Major Occupational Group, an NCS job classification.[14] We then use a crosswalk (created by OPM) to match occupational codes in the CPDF to MOGs in the NCS.[15] We then match each worker by MOG, grade, and year to the nearest corresponding NCS observation located within 15 miles.[16]

This process matches about half of the CPDF worker-year observations to the NCS. Observations are not matched if, in a given year, there weren't any nearby NCS workers who had the same grade and MOG. Most of the unmatched observations are workers in the Rest of US locality, where cities and towns were too small to justify frequent sampling.

Extracting outside wage indices from the NCS. Surveyors were given explicit instructions about how to assign pay grades to jobs. While it seems likely that they grading standards were applied to surveyed workers in a consistent manner, we want to avoid imputing too much meaning into these imaginary grades. Also, the NCS samples were relatively small and infrequent. Most MOG-by-locality cells were only sampled once every two or three years. Hence, we also want to avoid imputing too much meaning behind year-to-year variation in the measured wages. Hence, we use a statistical decomposition that extracts a time- and grade-invariant component of each outside wages. Let $j = 1, ..., J$ index each unique MOG-by-location cell, let i index a worker, let t index time, and let $g = 4, ..., 15$ index a GS pay grade.

$$\ln wo_{i,j,g,t} = \alpha_j + \psi_g + \kappa_t + \epsilon_{i,t}, \qquad \mathbb{E}\epsilon_{i,t} = 0$$

where $\ln wo_{i,j,t}$ is the log outside wage from the matched NCS data. To construct our

[14]In most cases, wages are annualized. In some cases, wages are reported at the hourly level, in which case we compute the annual wage assuming 2,000 annual hours.

[15]Our mapping requires some intermediate details. It is available upon request.

[16]Specifically, we obtain the latitudes and longitudes of the worker's office and of the survey area's center. We calculate the distance between the locations using the Haversine formula (see https://en.wikipedia.org/wiki/Haversine_formula)

outside wage index, we estimate the above equation via OLS, and compute the predicted log outside wage $\widehat{\ln wo_j}$ upon substituting $t = 1994$ and $g = 10$ into the estimated equation.

Sample selections. We use two samples of the CPDF data in this analysis. The *full* sample is used for basic summary statistics and simulation exercises. Subsample 1 covers workers who begin their service prior to the start of our dataset and is used to structurally estimate our quitting model in Section 3.4.

For the full sample, we drop individuals that are ever observed with one of the following conditions: leaving and then returning to the GS pay system while remaining in the CPDF, being employed at a grade level below 4 (considered blue-collar), or working in a non-full-time capacity. We drop individuals whose location has ever been omitted[17], and we drop a small number of individuals who had seemingly inconsistent information over time. From these observations, about 13% are missing a year or more of intermediate data during their years of service. Some of these individuals appear to have quit and then returned to service (a path explicitly ruled out by our model). However, more than three-quarters have no evidence of a quit preceding the gap of data. Hence, we suspect that the data for most of these individuals is truly missing, perhaps because they switched into the Department of Defense or Postal Service. All individuals with a gap in the data are dropped. This leaves us with about one million unique workers and five-million worker-year observations. We also drop a very small number of worker-year observations with age bins of 70-75 or 76-80.

Subsample 1 is used for estimating state-to-state transition rules, and is a subset of the full sample with the following additional conditions. Because our model rules out the possibility that workers endogenously switch locations (otherwise the relevant state-space would dramatically expand). However, we drop the roughly ten percent of remaining workers who violate this at some point in their careers.

Subsample 2 subsets further. It is used to estimate the preference parameters in the quitting model via maximum likelihood, which requires observing each worker's full career. Fur-

[17]Location information was redacted by OPM for individuals with occupations that were deemed "sensitive," including for example Secret Service agents, and agents of the Department of Alcohol, Tobacco, and Firearms.

thermore, to avoid making complicated assumptions about how workers anticipated FEPCA, we only wish to use data from 1994 and after. Therefore, subsample 2 only uses workers who entered the data-set with less than a year of creditable federal service, in the year 1994 and after. This restriction keeps about eighteen percent of the remaining sample, or 94,000 individual workers. To reduce the dimensionality of the state space, we then drop individuals who started federal service with an age bin suggesting that they were older than 30 at the time they started federal service, leaving us with about 54,000 workers.[18]

3.4 Estimating the Dynamic Retention Model

In this section we will introduce some modeling assumptions specific to our setting and then present our estimates of internal labor supply using the CPDF data. The compensation policy is treated as invariant during the sample period. The workforce is split into a grid of job-cells $j = 1, ..., J$ based on their locality's estimated locality pay growth rate, and the outside wage index of their locality-occupation pair using the MOG occupational classification described in Section 3.3.

3.4.1 The Inside Wage Process

The period wage $w_{i,t}$ consists of two components. The first component is basic pay, $wb_{i,t}$, which corresponds to the shock $u_{i,t}$ in the general framework in Section 3.2. Basic pay is presumed to be correlated in an unknown way with outside options. The other component is the locality pay rate, $F_{i,t}$. Total pay is given by

$$w_{i,t} = \left(1 + F_{j(i),t}\right) wb_{i,t}. \tag{3.5}$$

[18]This implies that years of experience and age essentially move one-to-one with each other and allows us to drop a state variable during the maximum likelihood estimation routine.

Basic pay. Basic pay evolves according to an AR(1) process:

$$\ln wb_{i,t+1} = \mu + \rho \ln wb_{i,t} + \sigma_b e_{b,i,t+1}, \tag{3.6}$$

where $e_{i,t}$ is i.i.d. standard normal.

Locality pay. The locality pay multipliers $F_{j,t}$ are assumed to follow the process

$$\ln\left(1 + F_{j,t+1}\right) = g_j + \ln\left(1 + F_{j,t}\right) + \sigma_F e_{F,j,t+1} \tag{3.7}$$

where $e_{j,t}$ is i.i.d. and mean-zero, and σ_F is perceived by workers to equal 0.[19] For the 48 federal localities observed during the span of our data, the implied annual growth rates range from 0.006 (Rest of US) to 0.017 (San Jose-San Francisco), with an unweighted average of about 1%.

To estimate the growth factors g_j, we used Subsample 1 of the data and regressed the left-hand-side of Equation 3.7 on a constant and linear time trend with coefficients specific to each locality. The fit of the model is excellent, with an R-squared above 0.989–hence, our simplifying assumption that workers believe $\sigma_{e,j} = 0$ appears quite reasonable. The estimated growth rates, along with 95% confidence intervals, are depicted in Figure 3.6.

Pension benefits. In addition to basic pay and locality pay, workers accumulate retirement benefits during their federal service. Most studies of federal civil servant compensation have concluded that the most important difference in fringe benefits between white collar employees in the federal government versus those in the private sector pertains to retirement benefits.[20] Hence, we explicitly model retirement benefits in the federal government.

Workers can access their accumulated benefits upon retiring, which usually occurs at age 65, regardless of whether they end their career in federal service. When a worker quits (or retires from) federal service in the model, we calculate the net present value of pension

[19] This simplification is to avoid having to integrate over a set of possible future locality pay factors when computing the expected future value function.

[20] See Falk (2012).

benefits and add it to her total outside option payoff. To calculate net present pension benefits, we use a formula provided by the General Schedule to compute annual pension payouts as a function of the worker's career history, which are paid every year between retirement and death and which are adjusted for inflation. The formula is a multiple of the average of the worker's top three annual wages (inclusive of locality pay), is increasing in federal government experience, and depends on whether the worker is subject to the old CSRS retirement system or the new FERS system (but all workers in our estimation sample our subject to the latter). Workers can claim the calculated pension benefit regardless of their employment status upon retirement. We use the FERS formulas on OPMs website and, for computation's sake, replace the top-three wage average with the top wage to reduce the number of relevant state variables in the dynamic optimization problem.

To account for the fact that pension benefits cease upon the worker's death, we integrate pension benefits over calibrated survival probabilities that decline linearly from 1 at age 65 to 0 at age 100.[21] The annual payout is $m(EXP_{i,t}) \times w_{i,t}$, which converts into time T net-present value

$$R_{i,t} = R(EXP_{i,t}, w_{i,t}) = \sum_{\tau=T}^{\infty} \pi(\tau) \times m(EXP_{i,t}) \times w_{i,t}. \tag{3.8}$$

There are several other differences between CSRS and FERS. As part of the CSRS system, mandatory retirement contributions equaling 7-8% of gross pay were deducted from worker's paychecks. Although the FERS system pays more generous pension benefits, workers don't pay into the system. FERS workers also receive an automatic one percent of their gross pay invested in a Thrift Savings Plan (similar to a 401K plan), as well as one-to-one matched contributions on the first three percent contributed and one-to-two matching on the next two percent. To capture these differences, we assume that a CSRS worker's effective wage is 92.5% of her gross pay while a FERS worker's effective wage is 104% of her gross pay.

Also, workers neither pay into nor accumulate social security benefits while in the CSRS

[21]The formula is $\pi(AGE) = (100 - AGE)/(100 - 65)$, yielding average remaining years of 17.5. For comparison, the 1994 US life tables estimated a life expectancy for the average 65 year-old of 17.4. See page 5 of https://www.cdc.gov/nchs/data/lifetables/life94_2.pdf.

system, whereas they *do* participate in social security under the FERS system. We assume that payments into and receipts out of social security are a wash as far as individual workers are concerned, and thus can be omitted from the payoffs of both inside and outside sector work. We don't attempt to account explicitly for the tax system. As long as workers pay an approximately constant fraction of their income into taxes over time and across states, then tax deductions will be implicitly captured in our estimate of β.

3.4.2 The Outside Option Process

Earnings after quitting. Workers are completely unobserved once they leave the government. Nevertheless, we put some theoretical structure into the outside option specification. Upon quitting (or retiring) at the end of year t, a worker starts earning outside wages $w_{i,t,\tau}$ in periods $\tau = t+1, t+2, ...$ up to and including the year of her retirement at age 65. In the year after retirement, she cashes in her federal retirement benefit $R(EXP_{i,t}, w_{i,t})$.[22] The net present value of quitting is therefore

$$V^o_{i,t+1} = \sum_{\tau=t}^{65-AGE_{i,t}} \delta^{\tau-t} wo_{i,t,\tau} + \delta^{65-AGE_{i,t}} R(EXP_{i,t}, w_{i,t}). \tag{3.9}$$

The outside wage $wo_{i,t,\tau}$. A worker's outside wage is a latent variable that is partially correlated with the worker's current basic pay, $wb_{i,t}$, a *relative* outside wage index, $\widehat{\ln wo_i}$, and an individual random effect v_i. In the first year after quitting, $t = \tau$, the outside wage is

$$\ln wo_{i,t,\tau} = \phi_0 + \phi^1 \ln wb_{i,t} + \phi_2 \widehat{\ln wo_i} + \sigma_{re} v_i. \tag{3.10}$$

After $t = \tau$, the outside wage grows at a constant rate such that $\ln wo_{i,t,\tau+1} = g_o \widehat{\ln wo_i}$. Hence, Equation 3.9 becomes

[22] Other potential retirement benefits are not explicitly unmodeled.

$$
\begin{aligned}
V^o_{i,t+1} =& V^o(AGE_{i,t}, EXP_{i,t}, wb_{i,t}, \widehat{\ln wo_{i,t}}, F_{i,t}, v_i) && (3.11)\\
=& \frac{1-(\delta g_o)^{65-AGE_{i,t}}}{1-\delta g_o}\left(\phi_0 + \phi_2 wb_{i,t} + +\phi_1 \widehat{\ln wo_i} + \sigma_{re} v_i\right) && (3.12)\\
&+ \delta^{65-AGE_{i,t}} R(EXP_{i,t}, wb_{i,t} * (1+F_{i,t})), && (3.13)
\end{aligned}
$$

where we remind readers that $V^o_{i,t+1}$ is relevant to the quitting decision made *at time* t–hence the explicit dependence on time t variables. The most important feature of this specification is that it allows basic pay to be correlated with both the outside option and the inside option in an economically plausible way. Importantly, we also allow for time-invariant unobserved heterogeneity in outside wages, captured by v_i. This is the usual practice of DRMs such as Gotz and McCall (1983) and Daula and Moffitt (1995), and can have very important implications for compensation policy.

Exogenous variation in locality pay. The main reason we include the outside wage index $\widehat{\ln wo_i}$ in Equation 3.9 rather than including it in the worker fixed effect v_i is to ensure the credibility of the assumption that the trajectory of locality pay is exogenous to v_i. Since locality pay was designed to reduce gaps between workers' pay and their local labor markets, it seems likely that the growth rates g_F would be correlated with these indices. However, conditional on the index, any remaining variation in locality pay trajectories is driven by the plausibly exogenous variation of the composition of nearby federal workers. Among two secretaries who are equally well-paid relative to their local markets, the one who happens to work alongside more technical federal employees will tend to earn more locality pay. To illustrate the quasi-random nature of locality pay, Figure 3.7 presents a scatterplot of the estimated MOG-by-locality fixed effects, $\widehat{\ln wo_{M,L}}$, for the log outside wage measure, $\widehat{\ln wo_{m,l}}$. The plot illustrates that many groups of employees had very similar relative local labor market indices, but very different locality pay trajectories.

The Bellman Equation. Under the above modeling assumptions, Equation 3.2 becomes

$$\bar{V}(AGE_{i,t}, EXP_{i,t}, F_{i,t}, wb_{i,t}, wo_{i,t}, v_i, g_{j(i)}) = \beta \times wb_{i,t}\left(1 + F_{i,t}\right) + \gamma$$

$$\delta \log\left(e^{\mathbb{E}_{i,t}\left[\bar{V}(AGE_{i,t}+1, EXP_{i,t}+1, F_{i,t+1}, wb_{i,t+1}, wo_{i,t+1})\right]} + e^{V^{o}(AGE_{i,t}, EXP_{i,t}, wb_{i,t}, wo_{i,t}, F_{i,t}, v_i)}\right), \tag{3.14}$$

where $V^{o}(\cdot)$ was provided in Equation 3.11.

3.4.3 Estimation of State-to-State Transitions

The stochastic transition rules for the state variables determining inside and outside payoffs can be estimated without solving the worker's optimal quitting problem. First, we estimate the AR(1) process for basic pay in Equation 3.6 using OLS. The estimates are included in Table 3.6. To check the normality assumption, the residuals from Equation 3.6 are normalized by the estimated standard deviations $\hat{\sigma}_b$ and then plotted against a standard normal distribution in Figure 3.6.[23]

Second, we fit annual data on locality pay multipliers to get an annual logarithmic growth rate for each locality, $g_j, j = 1, ..., J$, giving equal weight to each year.

3.4.4 Estimating Remaining Parameters

We calibrate the discount factor $\delta = 0.90$.[24] The remaining parameters to be estimated are $\theta = \left(\beta, \phi_0, \phi_1, \phi_2, \{g_j\}_{j=1}^{J}, \sigma_{re}\right)$. Despite our economically motivated specification for outside

[23] If stay/quit decisions are endogenous to future pay, then workers anticipating less favorable shocks v_i would select out of the sample and the estimated wage process will be biased. Although our model setup rules this out by assuming that workers observe next period's wage only after deciding to stay, it would still be ideal to estimate a wage process that is robust to this form of sample selection. One solution would be the Two-Step Heckman Correction procedure, which would require an instrumental variable for the retention probability that is exogenous to the future pay. If we found that individuals with a lower estimated propensity to stay in the government tended to draw higher future pay (conditional on current pay), than this would be indicative of a clear problem. We estimated a sample selection model using locality pay growth rate as instrument, and didn't find any evidence of this problem.

[24] Most of the dynamic retention literature has used, or estimated, discount factors close to 0.9. However, Warner and Pleeter (2001) estimated discount factors for retiring members of the military as low as 80%.

options, we do not observe these outcomes directly. Instead, they are estimated indirectly via revealed preference, i.e., quit rates, and are estimated via full-solution maximum likelihood estimation of the quitting model.

The likelihood function. We can write down a straightforward likelihood function implied by the quitting model, but it needs to be adapted to account for latent unobserved variables. First, recall from Section 3.3 that instead of observing true age, we observe the worker's starting age after it has been injected with error and then binned. Using Bayes' Theorem, we can recover a posterior distribution over the worker's true starting age. Each age bin has five possible integer ages, so let A denote the midpoint age (with $A-2$ and $A+2$ the lower and upper bounds) and let a denote the true starting age.

$$Pr(a|A) = \frac{Pr(a)Pr(A|a)}{\sum_{\tilde{a}=A-2}^{A+2} Pr(\tilde{a})Pr(A|\tilde{a})}.$$

We make the simple assumption that the probability mass function for starting age, $Pr(a)$, declines linearly from age 22 down to 0 at age 65. I.e.,

$$f(AGE) = \begin{cases} c(65 - AGE) \text{ if } 22 \leq AGE < 65 \\ 0 \text{ otherwise }, \end{cases} \tag{3.15}$$

where c is a constant of proportionality ensuring that integration to one. This assumption is sufficient to determine $Pr(a|A)$ above. To check the quality of the assumption, we compare the implied distribution over the error-injected and binned age variable to that which is observed empirically in Figure 3.8. The fit implied by our simple assumption appears to be quite good.

$Pr(A|a)$ can be recovered from

$$Pr(A|a) = Pr(a+e_a \in [A-2, A+2]|a) = \frac{1}{5}||\{a-2,\dots,a+2\} \cap \{A-2,\dots,A+2\}||. \tag{3.16}$$

The assumed prior probability mass function for a, the implied distribution of A, and the implied posterior distribution of a conditional on A are all graphed in Figure 3.9. Meanwhile, the vs are i.i.d. standard normal by assumption. Conditional on (a, v) and individual i's state variables at time t, the quitting probability $p_{i,t}(a, v)$ satisfies

$$p_{i,t}\left[a,v\right](\theta) = \frac{1}{1+\exp(o_{i,t}[a,v](\theta))}, \tag{3.17}$$

$$o_{i,t}\left[a,v\right](\theta) = V^o(AGE_{i,t}+1-a, EXP_{i,t}+1, wb_{i,t}, \hat{wo}_{i,t}, F_{i,t}; \theta, v) \tag{3.18}$$

$$-\mathbb{E}_t\left[V\left(AGE_{i,t+1}-a, EXP_{i,t+1}, wb_{i,t+1}, \hat{wo}_{i,t+1}, F_{i,t+1}; \theta, g_{j(i)}, v\right)\right]. \tag{3.19}$$

Time is normalized to run from years $t = 1$ through $t = T$, and $s_{i,t}$ captures whether the individual is in the data at time t. Hence, if an individual quits in year t, it means that $s_{i,t} = 1$ and $s_{i,t+1} = 0$.

The individual's entire likelihood function, conditional on (a, v), is

$$l_i\left[a,v\right](\theta) = \prod_t^T \left(s_{i,t+1}^{1-p_{i,t}[a,v](\theta)}(1-s_{i,t+1})^{p_{i,t}[a,v](\theta)}\right)^{s_{i,t}}.$$

Since neither a nor v are observed, we must integrate them out and work with the unconditional likelihood.

$$l_i(\theta) = \int\int l_i\left[a,v\right](\theta)dF_{a,v}(a,v) = \sum_{a=A_i-4}^{A_i+4} Pr(a|A_i)\int l_i\left[a,sv\right](\theta)\phi(v),$$

where we are taking advantage of the fact that $Pr(a|A_i) = 0$ if $|a - A_i| > 4$. Our estimates are chosen to maximize the sample log likelihood:

$$\hat{\theta} = \arg\max_\theta ll(\theta) = \arg\max \sum_i \ln l_i(\theta). \tag{3.20}$$

3.4.5 Computational Implementation and Results

We use a simple quadrature approximation of the integral over the normally distributed random effects by partitioning the unit interval into K segments of equal length, taking the midpoints of each segment $\{v_1, ..., v_K\}$, and rewriting the integral in Equation 3.20 as

$$l_i(\theta) \approx \sum_{a=A_i-4}^{A_i+4} Pr(a|A_i) \frac{1}{K} \sum_{k=1}^{K} l_i \left[a, v_k\right](\theta).$$

Calculating the likelihood requires solving the value function via backward induction. We explicitly solve for the value function along a set of discrete grid-points. We use linear interpolation to approximate off-grid values when demanded by backward recursion.[25] To approximate the expectation of the value function taken with respect to the innovations to basic pay, $u_{i,t}$, we use the same quadrature approximation used for the individual random effects. The estimated parameters are included in Table 3.7. To estimate the asymptotic variance of the estimates, we computed the observed Fisher information associated with the sample.[26]

The numeraire in the model is the standard deviation of the i.i.d. Gumbel preference shocks.[27] β reflects the value of net present earnings relative to the size of the shocks, and it also has a clear *pay elasticity* interpretation that relates back to previous studies. Using Equation 3.3, we can surmise that a one-time unanticipated increase in next period's pay, that neither affected the distributions of pay or outside options in other periods, would reduce the log-odds of quitting by 0.006 for every \$1,000. For a typical young worker earning \$45,000, with a quit rate of 6%, this translates into a pay elasticity of 25.4%.[28]

The estimated random-effects' standard deviation of $\hat{\sigma}_{re} = 0.8891$ indicates that time-invariant unobserved heterogeneity is very important relative to compensation. Consider a newly hired 25-year-old worker earning \$45,000 in basic pay having the typical outside-wage index of 4.3 (for simplicity we shut down locality pay in this estimate). Her quit probability is estimated to vary from 0.3% to 45.9% depending on whether her random effect is one standard-deviation below or above the mean. This implies that workers are predicted to

[25] See the package repositiory at https://github.com/JuliaMath/Interpolations.jl.

[26] We calculated the Hessian of the sample log-likelihood function analytically using forward-mode automatic differentiation. See the package repository at https://github.com/JuliaMath/ForwardDiff.jl.

[27] To be precise, the standard Gumbel distribution has a standard deviation equal to $\pi/\sqrt{6} = 1.23$.

[28] However, committing to a wage policy that pays older workers well may still be desirable so as to induce retention of young workers. We'll revisit this in the counterfactual simulations.

rapidly self-select themselves over time, and has strong potential implications for optimal compensation policy. Workforces comprised of employees with low tastes for outside work are cheaper to maintain. Compensation policies with low initial wages and steep earnings growth, or benefits that vest with tenure, will most likely to do a better job at achieving this favorable distribution of employee tastes.[29]

We estimate ϕ^1 to be negative, which implies that workers who receive higher base pay tend to have unobservedly lower outside options. Thus, exogenous policy-induced increases in compensation are associated with smaller increases in retention than endogenous increases. We estimate ϕ^2 to be positive, implying, unsurprisingly, that job-cells associated with larger measured outside-wage premia tend to exhibit higher quit rates, ceteris paribus.

3.5 Conclusion

In this paper, we estimated a dynamic retention model using exogenous variation in net-present compensation induced by FEPCA. Despite the relatively low quit rate in the federal government, retention remains fairly elastic to pay, with a pay elasticity of about 25% for a typical young worker. Workers who earn above average inside pay are estimated to have unobservedly *lower* outside options. Our model also suggests the existence of large time-invariant individual heterogeneity in the preference for government work, creating a *husbandry* mechanism where different pay policies can alter not only the retention rate, but also the type of workers who are retained.

Future work and extensions. We will use the dynamic retention model estimated in this paper, combined with auxiliary assumptions about counterfactual government hiring, to predict the impact of FEPCA and alternative compensation policies on the total costs of managing the civil servant workforce.

One of the major simplifying assumptions of our approach is that recruitment is exoge-

[29]However, our framework does not allow us to make conclusions about unobserved heterogeneity in worker quality. If workers with persistent unobservedly high preference for outside work are also more productive, then this would work against these cost-based conclusions.

nous to compensation policy. However, one might expect that more favorable compensation policy will result in a different composition of workers applying for and being accepted into federal jobs. Zhang and de Figueiredo (2018) recently find evidence that federal employees hired during recessions tend to earn large and persistent wage premia compared to federal workers hired in normal times, suggesting that the government succeeds in recruiting more competent employees when it becomes more attractive relative to the private sector. Thus, endogenizing the recruitment process to depend on compensation policy may give more plausible results.

3.6 Omitted Proofs

Proof of Lemma 3. The cost of retaining the marginal worker is paying her a wage that makes her indifferent to either quitting or accepting her best offer from the outside firms. As soon as the incumbent is observed making the equilibrium counteroffer, it is revealed that $z \geq Z$. Then the incumbent observes the shock ε and has the opportunity to adjust its offer by ε. By quitting, Claim 1 says that the worker's posterior ability becomes Z with probability one. I will now show that, after a counteroffer is received, quitting to the unattached market and receiving $S_t(Z)$ is preferred to immediately taking an outside offer.

By taking the outside option, the worker gets continuation value $\tilde{V}$ equaling the continuation value of a worker who didn't receive a counter offer, *plus* a term reflecting the future benefit of having posterior ability Z rather than $(z_{1,t}, Z)$. Let θ' denote the firm making the outside offer.

$$\tilde{V} = V_t(z_{1,t}, Z) + \delta \left(S_{t+1}(Z) - (\lambda V_{t+1}(\theta', z_{1,t}, Z) + (1-\lambda) V_{t+1}(z_{1,t}, Z)) \right).$$

Hence,

$$\begin{aligned} S_t(Z) - \tilde{V} &= s^\star(Z) + \delta \left(\lambda V_{t+1}(\theta', z_{1,t}, Z) + (1-\lambda) V_{t+1}(z_{1,t}, Z) \right) - V_t(z_{1,t}, Z) \\ > \mathbb{E}_z \left(s(\theta', z) | z_{1,t} < z \leq Z \right) &+ \delta \left(\lambda V_{t+1}(\theta', z_{1,t}, Z) + (1-\lambda) V_{t+1}(z_{1,t}, Z) \right) - V_t(z_{1,t}, Z) = 0. \end{aligned} \tag{3.21}$$

Thus, the incumbent will guarantee continuation value $S_t(Z)$, the value of quitting, and nothing more, in order to retain the worker, which it does in the event that $z \geq Z$, which occurs with probability $p^R = \frac{1-Z}{z_{2,t} - z_{1,t}}$.

Let $Z = Z_t(\theta_t, z_{1,t}, z_{2,t})$. From Lemma 5, a retained worker must receive continuation value $S(Z)$. If she is retained, then she gets the counteroffer w^R. Then, with probability δ, she survives into next period, and enters attached to θ_t with probability λ, and unattached with probability $(1-\lambda)$, in both cases having posterior $(Z, z_{2,t})$. Taken together, this implies

$$S_t(Z) = w^R + \delta \left(\lambda v_{t+1}(\theta_t, Z, z_{2,t}) + (1-\lambda) V_{t+1}(Z, z_{2,t}) \right).$$

The result follows by isolating w^R. □

Proof of Proposition 1. Suppose $T = 0$. Then trivially the claims hold. Now, suppose the claim holds for the model where $T = K$, as long as the worker's posterior ability can be described as $z \sim \mathcal{U}(z_{1,t}, z_{2,t})$. Then I will prove that these claims hold for the version of the model where $T = K + 1$ and conclude by induction that they hold for all possible T.

In her first year of age $t = 1$, the worker's posterior ability is uniform $(z_{1,1}, z_{2,1})$. Suppose she is unattached. The value to a vacant firm of recruiting the worker is her expected surplus, plus the future profits associated with becoming the incumbent. In this case, the worker's posterior remains unchanged, and so the continuation game is described by the $K - 1$ period model. Thus, the maximum value across vacant firms is strictly a function of her posterior and Claim 2 holds.

Now consider the counteroffer decision and let us momentarily imagine that ε has a small but positive variance. The incumbent's counteroffer decision will signal something about the worker's ability. There must exist a minimum offer, w^R, that the worker is willing to accept. The incumbent will always pay this amount if it intends to retain the worker.

If the incumbent makes the counteroffer w^R, it must anticipate a positive chance of wanting to retain the worker. It then observes ε and decides whether or not to adjust the offer. The payoff of retaining the worker is $s(\theta, z) + \varepsilon - w^R$, which is increasing in z. It will never adjust the offer upwards, because paying w^R is sufficient to retain the worker. If it adjusts the offer downwards, then it is guaranteed to lose the worker. Hence, it will only adjust the offer when $s(\theta, z) - w^R \leq -\varepsilon$. Thus, if quits after counteroffers occur with positive probability, it must be occurring because the offer was adjusted downwards.

I am interested in the limiting case where the variance of ε goes to 0. In this case, we see that the offer can only be adjusted when $s(\theta, z) - w^R \leq 0$. Writing a counteroffer costs an infinitesimally small amount, so the incumbent will only write a counteroffer if doing so generates positive expected payoffs. As ε converges in probability to 0, the expected payoff of counteroffering w^R converges to $s(\theta, z) - w^R$, which is increasing in z. Thus, the incumbent uses a cutoff rule: write the initial counteroffer w^R if and only if $z \geq Z$. If the worker is then observed quitting, it must be that the offer was privately adjusted downward, and $z = Z$. This concludes the proof of Claim 1.

□

Proof of Proposition 2. The proof uses induction on T. First, the statement is trivially true for all $t > T$ (in which case each function equals zero). Now, suppose the statement is true for all $t \geq K+1$. I will prove that it is true for $t = K$ and thus complete the proof by induction. First, notice that because $s(\theta, z)$ is HD-ϕ (Assumption 1), the full-information match $\theta^\star(z)$ must be HD-1, and the full-information surplus $s^\star(z)$ is HD-ϕ. The cutoff rule is the unique value satisfying Equation 1.4. Suppose Z satisfies the equation under the state $(\theta_t, z_{1,t}, z_{2,t})$. Then rescale both sides by α^ϕ to get

$$\alpha^\phi s(\theta_t, Z) + \delta \left(\lambda \alpha^\phi V_{t+1}(Z, z_{2,t}) + (1-\lambda)\alpha^\phi v_{t+1}(\theta_{t+1}, Z, z_{2,t})\right) = \alpha^\phi S_t(Z). \tag{3.22}$$

The time $t+1$ functions and $s(\theta, z)$ are HD-ϕ by assumption, and I showed that $S_t(\cdot)$ is HD-ϕ. Hence, we get

$$s(\alpha\theta_t, \alpha Z) + \delta \left(\lambda V_{t+1}(\alpha\alpha Z, \alpha z_{2,t}) + (1-\lambda) v_{t+1}(\alpha\theta_{t+1}, \alpha\alpha Z, \alpha z_{2,t})\right) = S_t(\alpha Z), \tag{3.23}$$

which implies that αZ is the cutoff under state $(\alpha\theta_t, \alpha\alpha z_{1,t}, \alpha z_{2,t})$. The time-$t$ cutoff rule is HD-1. A worker's time-t retention probability conditional on the public history is

$$Pr^{retain}(\theta_1, z_{1,t}, z_{2,t}) = \frac{z_{2,t} - Z_t(\theta_t, z_{1,t}, z_{2,t}}{z_{2,t} - z_{1,t}},$$

which is clearly HD-0. Rescale the arguments of Equation 1.2 by α and without loss of generality replace the maximizer θ^v with $\alpha\theta^v$, and the integrand z with αz.

$$V_t\left(\alpha\alpha z_{1,t}, \alpha z_{2,t}\right) = \max_{\alpha\theta^v} \left\{\mathbb{E}_z\left(s(\alpha\theta^v, \alpha z) | \alpha\alpha z_{1,t} \leq \alpha z \leq \alpha z_{2,t}\right) + \lambda \Pi_{t+1}(\alpha\theta^v, \alpha z_{1,t}, \alpha z_{2,t})\right\}.$$

By Assumption 1, we know that the expectation in the first term is HD-ϕ. By assumption,

Π_{t+1} is HD-ϕ. Hence,

$$V_t\left(\alpha\alpha z_{1,t},\alpha z_{2,t}\right)=\max_{\theta^v}\alpha^{\phi}\left\{\mathbb{E}_z\left(s(\theta^v,z)|z_{1,t}\leq z\leq z_{2,t}\right)+\lambda\alpha^{\phi}\Pi_{t+1}(\theta^v,z_{1,t},z_{2,t})\right\}=\alpha^{\phi}V_t\left(z_{1,t},z_{2,t}\right),$$

which proves that the unattached value function is HD-ϕ.

Almost identical arguments allow us to use the aforementioned homogeneity properties to conclude that the unattached value function in Lemma 5 and the ex-post profit function in Lemma 6 are both also HD-ϕ. □

Proof of Proposition 5. There are many ways to justify this assumption. For example, in the infinite horizon case, a worker's long-run average wage converges almost surely to their full-information surplus. Alternatively, the econometrician may observe some other outcome that is correlated with ability, such as test scores. In both cases, a worker's average percentile in the distribituion conditional on their $\mathbf{x}$ reveals the median for each $\mathbf{x}_k$. On the other hand, there may be a binary outcome, such as obtaining a certification, that reveals whether ability is above a given percentile.

Assume without loss of generality that X is in descending order of ρ_k or p_k. The following lemma explains that ranking the $\mathbf{x}$s according to these statistics must produce a first-order stochastic dominance ranking.

Lemma 8. Let us say that $k\succeq j$ if $z_1(\mathbf{x}_k)\geq z_1(\mathbf{x}_j)$ and $z_2(\mathbf{x}_k)\geq z_2(\mathbf{x}_j)$, with $\succ$ if at least one is strict. Then $p_k\geq p_j$ or $\rho_k\geq\rho_j$ if and only if $k\succeq j$ (resp., $>$ and $\succ$).

Proof. Consider the element K with the largest value of ρ_k (or p_k, respectively). Then I will first claim that $z_2(\mathbf{x}_K)>z_2(\mathbf{x}_j)$, for all $j\neq K$. To see why, suppose not. Then let J denote the set of event(s) with the maximum upper-bound $\bar{z}_1=1$, and let l denote the event with the second-highest upper-bound, which may or may not equal $z_2(\mathbf{x}_K)$.

By the uniform assumption, we need $\frac{Pr(z_1(\mathbf{x}_K)<z\leq z_2(\mathbf{x}_K))}{Pr(\bar{z}_2<z<\bar{z})}=\frac{z_2(\mathbf{x}_K))-z_1(\mathbf{x}_K)}{\bar{z}-\bar{z}_2}$.

We know that $Pr(z_1(\mathbf{x}_K)<z\leq z_2(\mathbf{x}_K))\geq p_K+\sum_{j\in J}p_j\frac{z_2(\mathbf{x}_K)-z_1(\mathbf{x}_K)}{\bar{z}-z_1(\mathbf{x}_j)}$ (">" in the event that $l\neq L$). We know that $Pr(\bar{z}_2<z<\bar{z})=\sum_{j\in J}p_j\frac{\bar{z}_2-z_1(\mathbf{x}_K)}{\bar{z}-z_1(\mathbf{x}_j)}$, which implies $\frac{Pr(z_1(\mathbf{x}_K)<z\leq z_2(\mathbf{x}_K))}{Pr(\bar{z}_2<z<\bar{z})} > \frac{z_2(\mathbf{x}_K))-z_1(\mathbf{x}_K)}{\bar{z}-\bar{z}_2}$ and thus contradicts the uniform assumption.

□

Hence, the top-ranked event first-order stochastically dominates the other events, and we can immediately infer that it has the highest upper-bound. If there are multiple ties for the top event, then all such events must correspond to the exact same posterior.

The next corollary explains how we can solve for posterior ability for any event once we have determined its upper-bound.

Corollary 3. $z_1(\mathbf{x}_k) = \frac{q^k - z_2\mathbf{x}_k q}{1-q}$ and $z_1(\mathbf{x}_k) = \frac{(\rho_k-1)z_2(\mathbf{x}_k)+\rho}{\rho_k}$.

Proof. The first part follows from writing the qth quantile as

$$q = \frac{q_k - z_1(\mathbf{x}_K)}{z_2(\mathbf{x}_K) - z_1(\mathbf{x}_K)}$$

and solving for $z_1(\mathbf{x}_K)$.

The second part follows from writing the success probability $\rho_k = \frac{z_2(\mathbf{x}_k)-\rho}{z_2(\mathbf{x}_k)-z_1(\mathbf{x}_k)}$ and solving for $z_1(\mathbf{x}_k)$, □

Thus, one can immediately identify the posterior of the highest-ranked event using this formula, since $z_2(\mathbf{x}_K) = 1$. Next, consider the event with the second-highest upper-bound. Based on the same logic as the earlier lemma, we can infer that the only way the unconditional distribution of ability can be uniform is if the second-highest upper-bound lines up exactly with either $z_2(\mathbf{x}_K)$ or $z_1(\mathbf{x}_K)$. Otherwise, there would be too much mass just below the second-highest cutoff relative to the mass above the second-highest cutoff. We can then verify which case we are in by determining whether we have fully accounted for the mass that should fall between $z_1(\mathbf{x}_K)$ and $z_2(\mathbf{x}_K)$.

Corollary 4. Let $j < K$. If event $j+1$ has upper-bound $z_2(\mathbf{x}_K)$, then event j's upper-bound is either $z_2(\mathbf{x}_K)$, or $z_1((x)_K)$.

Event j's upper-bound equals $z_2(\mathbf{x}_K)$ if and only if

$$p_K + \sum_{k=j+1}^{K-1} p_k \frac{z_2(\mathbf{x}_K) - z_1(\mathbf{x}_k)}{z_2(\mathbf{x}_k) - z_1(\mathbf{x}_k^H)},$$

where $z_1(\mathbf{x}_k^H)$ is the lower-bound for event k using Corollary 3, under the assumption that it has upper-bound $z_2(\mathbf{x}_K)$.

This implies a recursive procedure that is guaranteed to identify every event-conditional posterior. First, identify the posterior of event K. Then, iteratively go down the list of events until hitting the first event J that cannot have the same upper-bound as K, according to Corollary 4. Store the lower-bounds of all the previous events implied by Corollary 3. Thus, we have identified the posteriors for all events greater than J and we have exhausted the mass that can be attributed to the interval $(z_1(\mathbf{x}_K), z_2(\mathbf{x}_K))$. Now, repeat the same procedure, conditioning on $z < z_1(\mathbf{x}_K)$. Continue until all the upper-bounds and lower-bounds have been identified.

□

Bibliography

Akerlof, George A. (1970), “The market for ”lemons”: Quality uncertainty and the market mechanism.” *The Quarterly Journal of Economics*, 84, 488–500.

Albouy, David and Mike Zabek (2016), “Housing inequality.” NBER Working Paper.

Altonji, Joseph G. and Charles R. Pierret (1998), “Employer learning and the signalling value of education.” In *Internal Labour Markets, Incentives and Employment*, 159–195, Palgrave Macmillan UK.

Altonji, Joseph G. and Charles R. Pierret (2001), “Employer learning and statistical discrimination.” *The Quarterly Journal of Economics*, 116, 313–350.

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

Asch, Beth J., James Hosek, and Michael G. Mattock (2013), *Model, Data, and Estimates*, 5–20. RAND Corporation.

Baker, George, Michael Gibbs, and Bengt Holmstrom (1994), “The wage policy of a firm.” *The Quarterly Journal of Economics*, 109, 921–955.

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.

Bernhardt, Dan (1995), “Strategic promotion and compensation.” *The Review of Economic Studies*, 62, 315–339.

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.

Bolton, Alexander, John M. De Figueiredo, and David Lewis (2018), “Elections, ideology, and turnover in the u.s. federal government.” *Academy of Management Proceedings*, 2018, 12717.

Borjas, George J. (1982), “Labor turnover in the u.s. federal bureaucracy.” *Journal of Public Economics.*

Braga, Breno (2018), “Earnings dynamics: The role of education throughout a worker’s career.” *Labour Economics*, 52, 83–97.

Bronson, Bronson & McKinnon (1980), “The law firm of bronson, bronson and mckinnon: Oral history transcripts and related material.”

Burdett, Kenneth and Dale T. Mortensen (1998), "Wage differentials, employer size, and unemployment." *International Economic Review*, 39, 257–273.

Carrell, Scott E (2007), "The national internal labor market encounters the local labor market: Effects on employee retention." *Labour Economics*, 14, 774–787.

Daula, Thomas and Robert Moffitt (1995), "Estimating dynamic models of quit behavior: The case of military reenlistment." *Journal of Labor Economics*, 13, 499–523.

DeVaro, Jed and Michael Waldman (2012), "The signaling role of promotions: Further theory and empirical evidence." *Journal of Labor Economics*, 30, 91–147.

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

Falk, Justin R. (2012), "Comparing benefits and total compensation between similar federal and private-sector workers." *The B.E. Journal of Economic Analysis & Policy*, 12.

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

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.

Fuchs, William and Andrzej Skrzypacz (2019), "Costs and benefits of dynamic trading in a lemons market." *Review of Economic Dynamics*, 33, 105–127. Fragmented Financial Markets.

Galanter, Marc and Thomas Palay (1993), *Tournament of Lawyers (paper)*. University of Chicago Press.

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. Theories of the Firm.

Garrison, Lloyd K. (1938), *The Economics of the Legal Profession*. American Bar Association.

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

Gotz, Glenn A. and John J. McCall (1983), "Sequential analysis of the stay/leave decision: U.s. air force officers." *Management Science*.

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

Greszler, Rachel and James Sherk (2016), "Why it is time to reform compensation for federal employees." *Heritage Foundation*.

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

Harbaugh, William Henry (1973), *Lawyer's Lawyer: The Life of John W. Davis*. Oxford University Press.

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.

Keane, Michael P. and Kenneth I. Wolpin (1997), "The career decisions of young men." *Journal of Political Economy*, 105, 473–522.

Knapp, David, Kristine Brown, James Hosek, Michael Mattock, and Beth Asch (2016), *Retirement Benefits and Teacher Retention: A Structural Modeling Approach*. RAND Corporation.

Krashinsky, Harry (2002), "Evidence on adverse selection and establishment size in the labor market." *ILR Review*, 56, 84–96.

Kreps, David M. and Robert Wilson (1982), "Sequential equilibria." *Econometrica*, 50, 863–894.

Lewis, Gregory B. and Samantha L. Durst (1995), "Will locality pay solve recruitment and retention problems in the federal civil service?" *Public Administration Review*, 55, 371–380.

Li, Jin (2013), "Job mobility, wage dispersion, and technological change: An asymmetric information perspective." *European Economic Review*, 60, 105–126.

Lise, Jeremy and Jean-Marc Robin (2017), "The macrodynamics of sorting between workers and firms." *The American Economic Review*, 107, 1104–1135.

Moscarini, Giuseppe and Fabien Postel-Vinay (2013), "Stochastic search equilibrium." *The Review of Economic Studies*, 80, 1545–1581.

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.

Rao, P. Poornachandra (1990), "A dynamic programming approach to determine optimal manpower recruitment policies." *The Journal of the Operational Research Society*, 41, 983–988.

Rudin, Walter R. (1976), *Principles of Mathematical Analysis*, 3 edition. McGraw–Hill.

Rust, John (1987), "Optimal replacement of gmc bus engines: An empirical model of harold zurcher." *Econometrica*.

Schonberg, Uta (2007), "Testing for asymmetric employer learning." *Journal of Labor Economics*, 25, 651–691.

Selten, R. (1975), "Reexamination of the perfectness concept for equilibrium points in extensive games." *International Journal of Game Theory*, 4, 25–55.

Spurr, Stephen J. (1990), "Sex discrimination in the legal profession: A study of promotion." *ILR Review*, 43, 406–417.

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

Utgoff, Kathleen Classen (1983), "Compensation levels and quit rates in the public sector." *The Journal of Human Resources*.

Volcker, Paul A (1989), "Leadership for america: Rebuilding the public service." *Report on the National Committee on the Public Service*.

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. SOLE/EALE conference issue 2015.

Warner, J. and M. Goldberg (1984), "The influence of non-pecuniary factors on labor supply: The case of navy enlisted personnel." *Review of Economics and Statistics*.

Warner, John T. and Saul Pleeter (2001), "The personal discount rate: Evidence from military downsizing programs." *American Economic Review*, 91, 33–53.

Zhang, Congshan and John M. de Figueiredo (2018), "Are recessions good for government hires? The effect of unemployment on public sector human capital." *Economics Letters*, 170, 1–5.

Figure 3.1: 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 3.2: Cardinal measures of law school quality

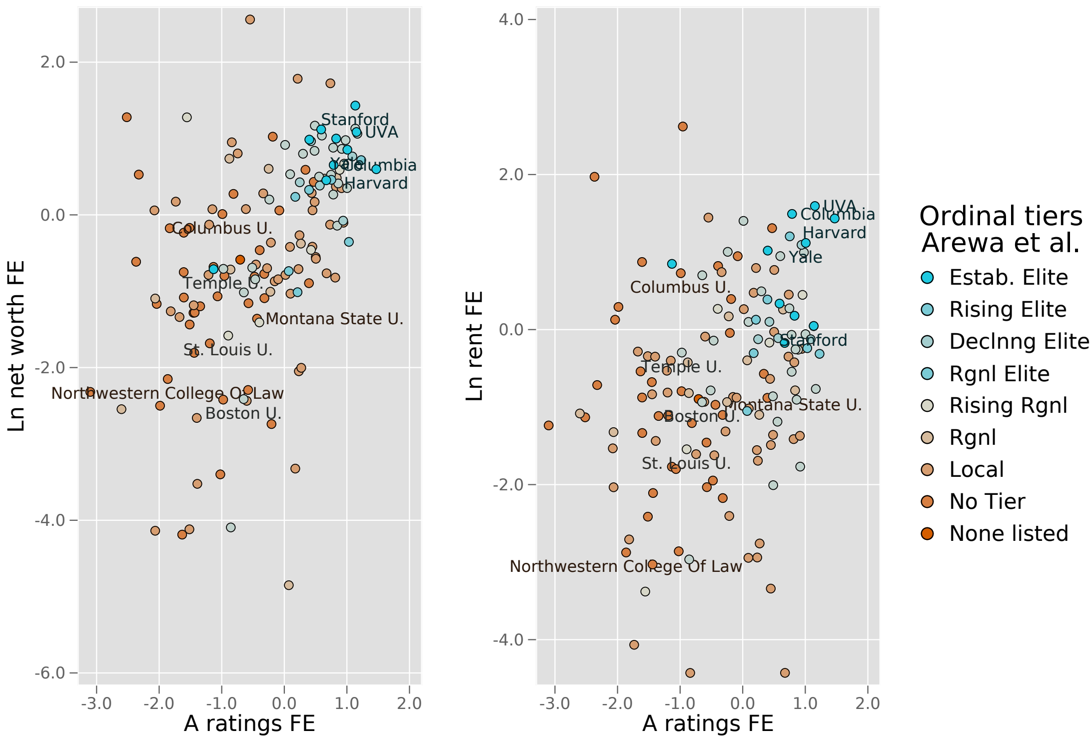


Figure 3.3: Overhead costs from Garrison (1938)

TABLE 28

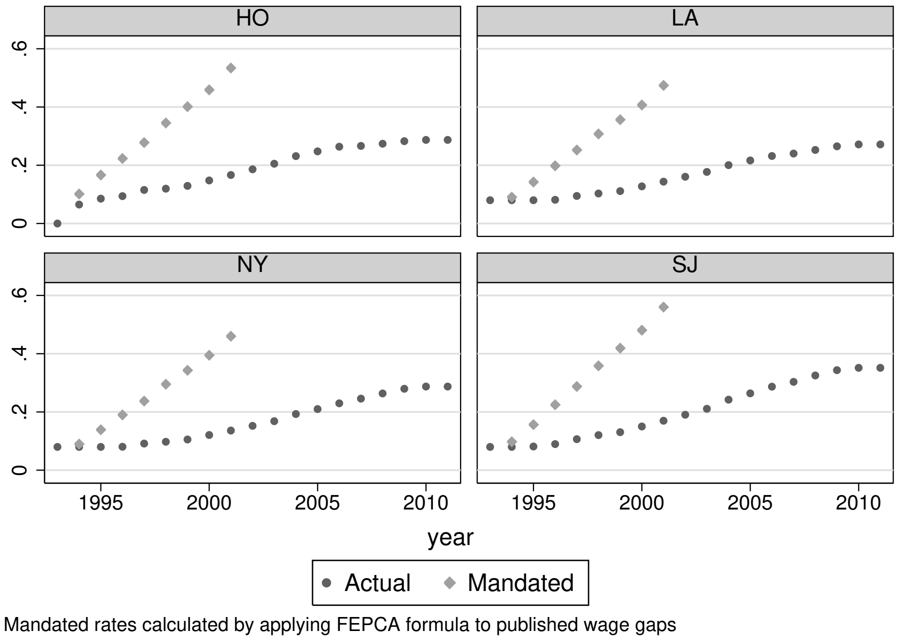
OVERHEAD EXPENSES OF 1,808 N. Y. COUNTY LAWYERS

(1933) Net professional income groups	Median Gross Income	Median Net Income	Median Expenses	Percentage, Median Expenses of median gross	(1928-32) Same percentages, based on 1928-32 incomes
Under $500	$ 688	$ 250	$ 438	63.7%	61.5%
$ 500– 999	1,292	750	542	41.9	34.7
$ 1,000– 1,499	1,773	1,250	523	29.6	29.1
$ 1,500– 1,999	2,412	1,750	662	27.4	28.0
$ 2,000– 2,499	3,297	2,250	1,047	31.8	30.9
$ 2,500– 2,999	3,898	2,750	1,148	29.3	27.2
$ 3,000– 4,999	5,291	4,000	1,291	24.4	23.4
$ 5,000– 7,499	8,701	6,250	2,451	28.2	28.0
$ 7,500– 9,999	11,987	8,750	3,237	27.0	28.4
$10,000–14,999	18,846	12,500	6,346	33.7	34.2
$15,000–24,999	30,625	20,000	10,625	34.7	36.9
$25,000–49,999	52,941	37,500	15,441	29.2	16.3
Whole group	4,389	2,939	1,450	35.3	26.2

81. *Ibid.*

82. New York County Report, *supra* note 1, p. 36, Tables 22 and 23.

Figure 3.4: Locality pay growth trajectories



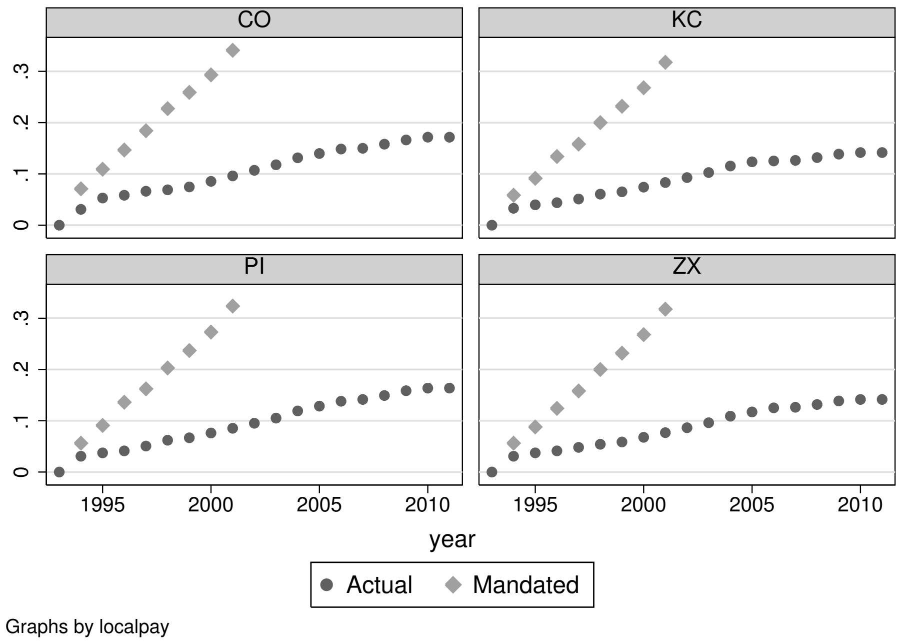


Figure 3.5: GS Pay Increases, 1965 – 2009

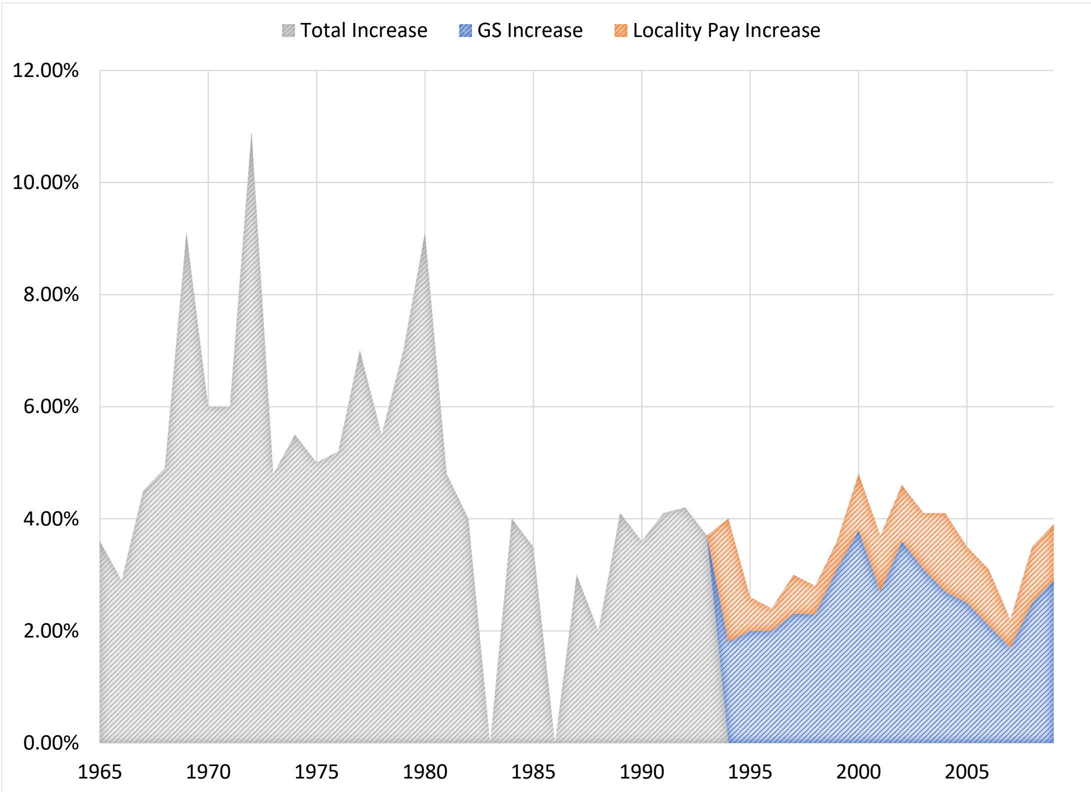


Figure 3.6: Estimated annual locality pay growth rates

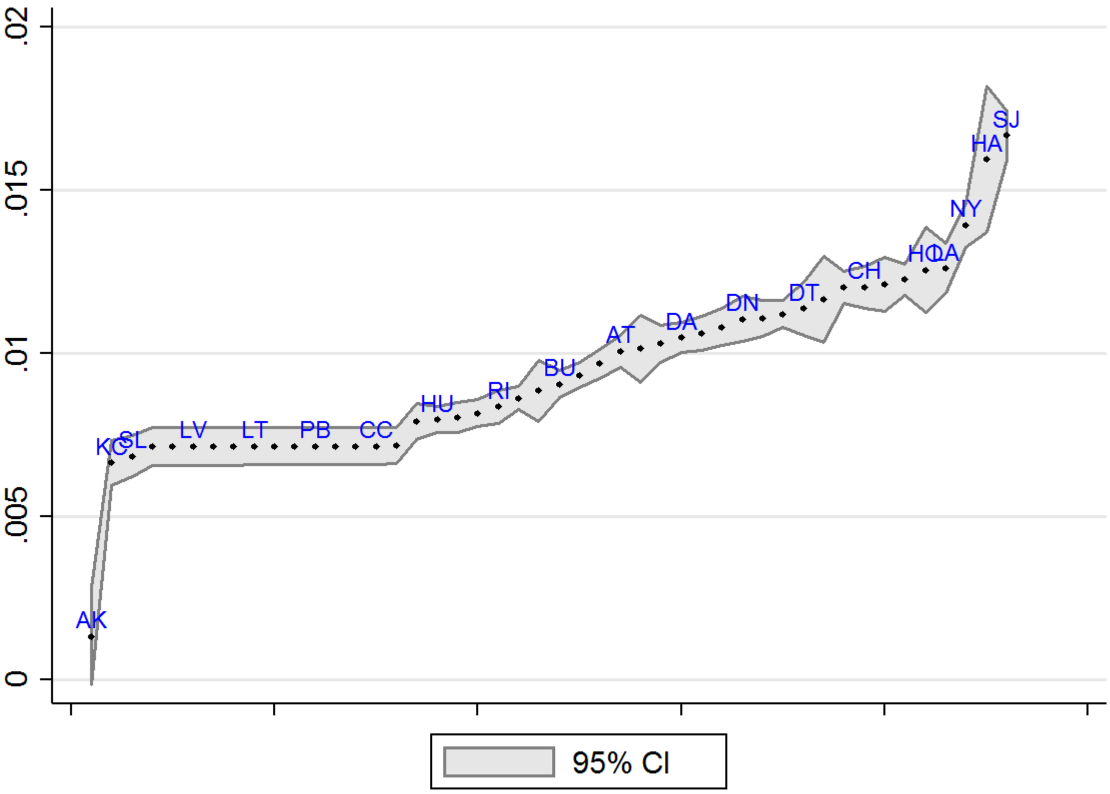
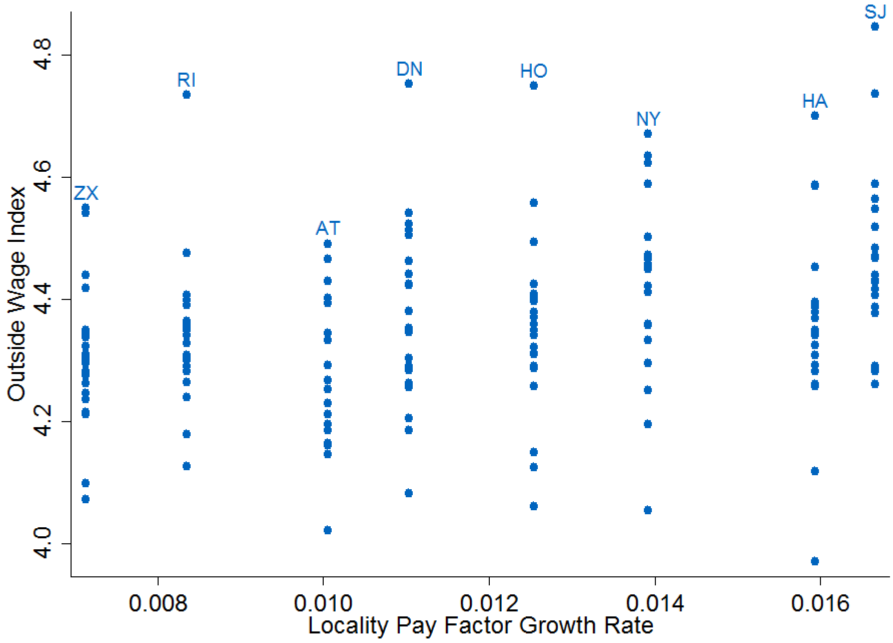


Figure 3.7: Occupation-by-Locality Outside Wages versus Locality Pay: Selected Localities



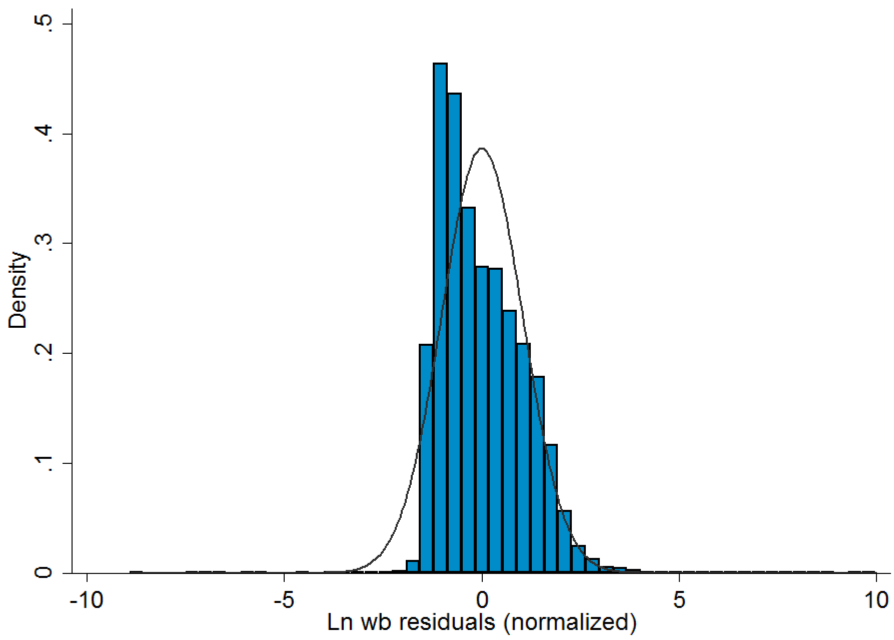
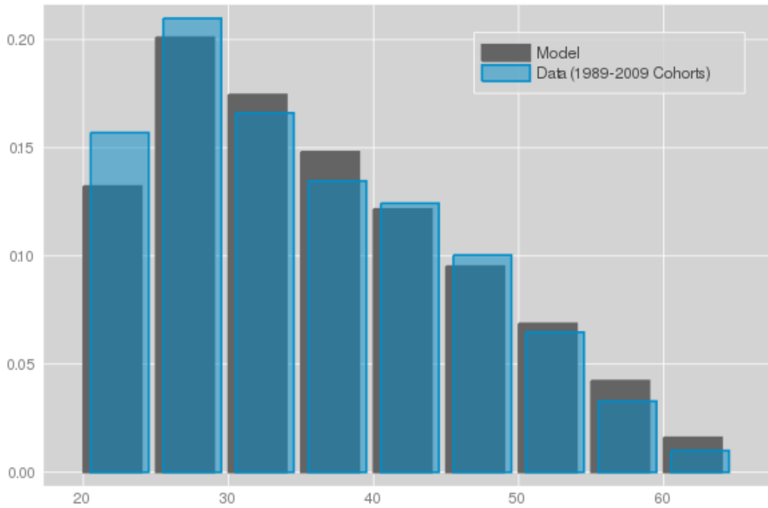


Figure 3.8: Comparison of Distributions over Error-Injected Binned Age Variable



.

Figure 3.9: Posterior distribution of age

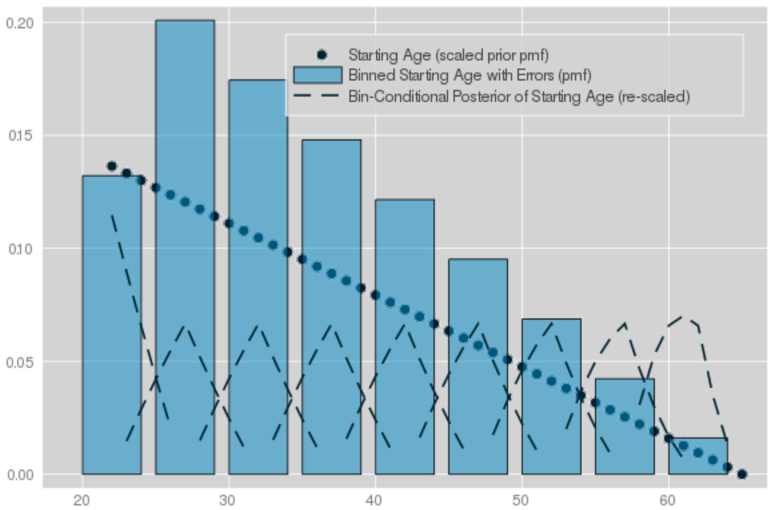


Table 3.1: Summary statistics by lawyer-year

	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
Separation (firm survives)	0.06			
Separation (firm exits)	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 3.2: 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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: 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 3.4: Change in firm rank

	(1)	(2)	(3)
Separation (firm survives)	-0.061***	-0.063***	-0.063***
	(0.001)	(0.001)	(0.001)
Separation (firm exits)	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
Omitted year is 1933
Std. errors in parenthesis * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: Linear Probability Models of Future a rating

	Obtains a rating		
	(1)	(2)	(3)
Separation (firm survives)	-0.048***	-0.045***	-0.041***
	(0.006)	(0.006)	(0.005)
Separation (firm exits)	-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 12+ 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 3.6: Estimated AR(1) Process for $\ln wb$ (000's)

Parameter	Estimate (Std. Error)
ρ	0.973***
	(7.19e-5)
μ	0.134***
	(2.88e-04)
σ	.047***
	(3.81e-05)
Person obs	277,079
Person-year obs	2,014,200

***Significant at 99% level.

Table 3.7: Parameters Estimated via Maximum Likelihood

Parameter	Estimate (Std. Error)
δ	0.90 (*calib.*)
σ_{re}	0.8891*** (.0469717)
β	0.00616831*** (0.000518685)
ϕ^0	4.5561*** (0.262226)
ϕ^1	-0.2393*** (0.0291)
ϕ^2	0.146676*** (0.0549141)
Person obs	54,120
Person-year obs	473,220

***Significant at 99% level.