NBER WORKING PAPER SERIES

ALLOCATING EFFORT AND TALENT IN PROFESSIONAL LABOR MARKETS

Gadi Barlevy
Derek Neal

Working Paper 23824
http://www.nber.org/papers/w23824

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2017

We thank John Kennan, Kevin Lang, Charles Lewis, Kevin Murphy, Canice Prendergast, Chris Taber, and Fabrice Tourre for useful discussions. We thank seminar participants at the University of Chicago, UCL, Edinburgh University, UW-Madison, the 2014 North American Summer Meetings of the Econometric Society, and the NBER Organizational Economics working group for valuable feedback. We thank Nicki Bazer, Ted Froum, Hank Kelly and Stephanie Scharf for useful conversations about personnel practices in law firms. We thank Natascha Geirhos and Henry Doorn for helpful insights concerning the public accounting industry. We thank Andrew Jordan, Caitlin McCarthy, Jorge Luis García, and Max Samels for excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Allocating Effort and Talent in Professional Labor Markets
Gadi Barlevy and Derek Neal
NBER Working Paper No. 23824
September 2017
JEL No. J01,J22,J44,M51

ABSTRACT

In many professional service firms, new associates work long hours while competing in up-or-out promotion contests. Our model explores why these firms require young professionals to take on heavy workloads while simultaneously facing significant risks of dismissal. We argue that the productivity of skilled partners in professional service firms (e.g. law, consulting, investment banking, and public accounting) is quite large relative to the productivity of their peers who are competent and experienced but not well-suited to the partner role. Therefore, these firms adopt personnel policies that facilitate the identification of new partners. In our model, both heavy workloads and up-or-out rules serve this purpose. Firms are able to identify more professionals who can function effectively as partners when they require new associates to perform more tasks. Further, when firms replace experienced associates with new workers, they gain the opportunity to identify talented professionals who will have long careers as partners. Both of these personnel practices are costly. However, when the gains from increasing the number of talented partners exceed these costs, firms employ both practices in tandem. We present evidence on life-cycle patterns of hours and earnings among lawyers that supports our claim that both heavy workloads and up-or-out rules are screening mechanisms.

Gadi Barlevy
Economic Research Department
Federal Reserve Bank of Chicago
230 South LaSalle
Chicago, IL 60604
gbarlevy@frbchi.org

Derek Neal
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and The Committee on Education
and also NBER
d-neal@uchicago.edu

Introduction

Many professional service firms employ two personnel practices that are uncommon in other labor markets. They ask young professionals to work exceptionally long hours, and they require them to compete in up-or-out promotion contests. Here, we develop a model that explains why.

Our work fills a hole in the existing literature on professional labor markets. Research on rat race models and models of career concerns provide reasons that young professionals may work long hours, but these literatures do not directly address why many professional service firms adopt up-or-out promotion rules. The literature on up-or-out explains how firms can use this policy to solve commitment problems or to facilitate the identification of talented professionals who will succeed as partners, but these models of commitment and screening typically ignore worker effort and make no clear predictions about the effort levels of young professionals relative to those of other young workers.

While heavy workloads and up-or out promotion rules may serve many purposes, our results suggest that these two practices are not separate phenomena. Both heavy workloads and up-or-out rules serve a common purpose. These practices facilitate the identification of the talented professionals who will lead their organizations in the future. Firms are able to identify more professionals who can function effectively as partners when they require new associates to perform more tasks, and when firms replace experienced associates with new workers, they gain opportunities to identify talented professionals who will have long careers as partners.

In our model, a new associate is participating in an audition for a leadership position. Throughout the paper, we use the title partner to denote this leadership position because, historically, partnerships have been a common organizational structure in professional service industries. However, the personnel practices that our model addresses continue to exist in many professional service firms that are no longer organized as partnerships.[1]

Professional service firms sell expertise to clients. So, whatever their legal structure, these firms cannot survive if they are not able to efficiently identify gifted leaders who are able to build and maintain relationships with valuable clients. We derive our key results in a model where auditions reveal which new associates possess the skills that partners require. We also present an alternative model where auditions reveal which new associates are able to acquire the skills that partners need. Many market mechanisms decentralize the planner's solutions to these models. We describe one mechanism and discuss its implications for how to interpret observed relationships between long hours and up-or-out rules.

In the penultimate section, we document several patterns in data on earnings and hours worked among lawyers that support our contention that professional service firms require new associates to take on heavy workloads while participating

[1]Some consulting firms, e.g. Accenture, and most investment banking firms are no longer organized as partnerships. Yet, young professionals in these firms continue to take on heavy work loads, and these firms continue to follow up-or-out promotion rules. In investment banking, the career ladder leads to managing director, and managing directors take on most of the roles that partners played when investment banks were partnerships. Modern investment banks are involved in many lines of business other than professional services. Our model is most applicable to groups of investment bankers who provide advice and services related to mergers and acquisitions. In these groups, up-or-out is the norm.

in up-or-out promotion contests because both policies speed the discovery of new partners. Our conclusion reviews our contribution and discusses future research that may shed more light on the evolution of personnel policies in professional markets.

1. Literature Review

New associates in elite professional service firms (e.g. leading firms in law, consulting, public accounting, and investment banking) often work much longer hours than most white collar workers who have similar levels of education. Building on the work of Akerlof [1976], Landers et al. [1996] offer a possible explanation for this pattern. They treat law firms as teams and assume that, while team output is observed, individual output is not. In their model, teams share profits according to the following rules: Partners pay associates a fixed salary and share remaining profits equally. Associates work schedules that partners dictate, and at the end of their terms as associates, they bid for shares in the partnership. Retiring partners sell their shares to the next generation of partners. Partners and associates have heterogeneous effort costs and also possess private information about these costs.

Partners benefit from hiring associates with low effort costs because this allows them to sell their equity shares to more productive lawyers in the future. In the separating equilibrium that Landers et al. [1996] describe, a menu of employment contracts specifies hours requirements and compensation for new associates in each law firm. Lawyers who share the same effort costs select the same contracts and work together in the same firms.

Ex post, associates in all firms work more than the efficient number of hours.[2] As in the canonical rat race model, Akerlof [1976], hours distortions are the equilibrium mechanisms that sort heterogeneous workers to heterogeneous teams. This model of hours requirements for new associates in professional labor markets does not directly address retention or promotion decisions. In the separating equilibrium that Landers et al. [1996] describe, no one leaves law as a profession, no one changes law firms, and all associates become partners.

Holmstrom [1999] provides a different reason that young professionals may work long hours. In his model, output is not contractible, so firms pay workers ex ante based on their reputation. Workers possess no private information about their ability levels. All market participants have the same prior beliefs about all workers and learn about all workers at the same rate by observing public output signals. However, workers do have private information about their effort levels, and Holmstrom [1999] shows that young workers may choose more than the efficient level of effort to increase the output signals that firms use to form beliefs about their abilities. In equilibrium, firms infer the workers' equilibrium effort choices and adjust their inferences about worker ability accordingly. However, as in rat race models, no individual worker has an incentive to deviate from the inefficient equilibrium.[3] Holmstrom [1999] does not address job assignments and therefore does not model promotion or retention decisions.

[2]Further, partners take on workloads that are below efficient levels since they share the returns of their efforts with other partners.

[3]In contrast to rat race models, equilibrium effort levels in career concerns models need not be excessive relative to efficient levels under full information. These models highlight one reason that work effort may decline over a worker's life cycle, but not all parameterizations yield the result that young workers begin their careers working too hard relative to efficient levels of effort.

Gicheva [2013] presents a model of variation in hours worked among young, well-educated workers. She notes that, if workers differ in their preferences for leisure, young workers who work the most are also most likely to enjoy promotions and high rates of wage growth in the future.[4] Gicheva [2013] offers her model as an explanation for the convex relationship between hours and earnings observed among well-educated workers. Goldin [2014] argues that the organization of work in certain environments drives this convex relationship. Returns to long hours can be great in jobs that require constant communication with customers and co-workers. Goldin [2014] cites private law firms as examples of organizations that often place a premium on long hours, but neither Gicheva [2013] nor Goldin [2014] address up-or-out rules.

The up-or-out literature contains several variations on two different approaches, but neither approach addresses why young professionals in up-or-out firms often work much more than other workers with similar levels of education. One literature characterizes up-or-out rules as commitment devices that solve a double moral hazard problem between workers and firms. Firms have private information about either the output of a worker or a worker's ability. Workers have private information about their actions. Firms want to provide workers with incentives to take efficient actions, but workers know that, ex post, firms may have an incentive to renege on payments linked to performance measures that only the firm observes. Firms solve this double moral hazard problem by making verifiable commitments to up-or-out promotion rules. Given these rules, if a firm were to make an unfavorable report about a worker who produced a positive output signal, the firm would expect to incur a loss. Thus, up-or-out rules allow firms to more credibly promise to reward hidden actions.[5]

This literature begins with Kahn and Huberman [1988], who argue that up-or-out allows firms to induce workers to make investments in firm-specific skills.[6] Prendergast [1993], Waldman [1990], and Ghosh and Waldman [2010] extend the logic of Kahn and Huberman [1988] to settings where various features of the employment relationship point to a specific reason that up-or-out rules may or may not make contingent promises concerning raises and promotions more credible.[7]

In Rebitzer and Taylor [2007], up-or-out rules help enforce a commitment to fix an overall ratio of associates to partners. This maintains partnership stability by

[4]This model extends the Gibbons and Waldman [1999] model of learning and assignment by allowing effort levels to affect rates of skill accumulation. Gibbons and Waldman [2006] incorporate formal schooling and task-specific human capital accumulation.

[5]Levin and Tadelis [2005] suggest partnerships solve a commitment problem between professional service firms and their customers. Professional service firms promise to supply talented professionals who perform quality work, but clients of professional service firms may find it difficult to judge the talent of different professionals ex ante. Revenue sharing within partnerships makes promises concerning the quality of professional services more credible because clients know that the other partners in the firm have agreed to share revenue with the partners who are directing their cases.

[6]See Gilson and Mnookin [1989] for an application of the Kahn and Huberman [1988] model to law firms.

[7]Ghosh and Waldman [2010] also model worker effort. However, in contrast to our model, worker actions are hidden in their model. Also, in their model, effort levels among new professionals may be high or low in both up-or-out firms and firms that follow standard promotion practices.

raising the cost of defection for any group of partners that consider taking their clients and forming a new firm.[8]

Both O'Flaherty and Siow [1992] and Demougin and Siow [1994] link up-or-out promotion rules to optimal screening procedures. Demougin and Siow [1994] do so in a model of hierarchies. In this model, firms decide what portion of their new workers they will train to be potential managers. This training may be interpreted as on-the-job learning or as a screening process that determines the suitability of workers for the management position. When the outside wage for new workers is high enough, all firms in a given industry choose to train or screen all new workers and dismiss all who are not deemed worthy of promotion. Those without the talent required to work as managers leave the industry, and if a given firm identifies more managers than it needs, these excess managers are hired away by firms that failed to identify enough managers.

Below, we derive results that link up-or-out rules with the option value logic found in both O'Flaherty and Siow [1992] and Demougin and Siow [1994]. Yet, our results differ in two ways. First, we model worker effort and introduce a signaling technology such that the market identifies more professionals who can function effectively as partners when new professionals perform more tasks. This allows us to explain why young professionals work long hours in the same sectors where up-or-out promotion rules are most common. Second, our comparative static results concerning when up-or-out regimes exist do not deal with changes in outside options but rather changes in the relative productivities of experienced professionals of different abilities who occupy different roles within the professional sector. Changes in technology or organizational structure that raise the relative productivity of experienced professionals who are skilled but not partner material make up-or-out less productive while changes that raise the relative productivity of partners make up-or-out rules more productive.

Some scholars argue that up-or-out rules are puzzling because they require the dismissal of workers who may be doing a competent job in their current positions. Our model demonstrates that, in up-or-out settings, associates never want to continue in their associate positions once they realize that they are not going to become partners. The long hours that associates work are a cost that young professionals pay to learn whether or not they are well-suited to become partners. If these young professionals learn they are not going to become partners, they are no longer willing to pay this cost. Further, young professionals take associate positions in order to learn whether or not they are capable of filing the partner role. Once they learn that they are not, their firms can no longer offer them this valuable learning opportunity, and we argue that, absent this opportunity, their most valuable employment options are typically elsewhere.

Finally, because we model workloads, our model provides new insights concerning life-cycle patterns of changes in hours worked among professionals who follow different career trajectories. We document that, among lawyers with roughly ten years of experience in private law firms, those who leave private law or leave the partnership track within private law reduce their hours significantly even though their wage rates often rise.

---

[8]Their model also implies that firms should limit the opportunities for associates to build relationships with clients.

This pattern is difficult to understand in most models of life-cycle labor supply. However, in our model, the heavy workloads that young associates tackle in private law firms are one component of an ordeal that young associates endure in order to audition for partnerships. When an associate learns that she is not going to make partner, her audition is over, and she may well reduce her hours even though her skill set continues to command a high wage rate.

2. Model Setup

Consider an environment with two sectors: a professional sector and an outside sector. In the outside sector, there is one job, and output does not vary with worker ability. There are two jobs in the professional sector, associate and partner, and output does vary with ability. We use these titles for convenience, but there are no productivity interactions between these positions, and associates do not work under a specific partner. We begin by describing worker preferences and the production technologies for each job.

2.1. **Preferences and Production.** Time is measured in discrete periods, and the time horizon is infinite. Each period, a unit mass of workers is born and lives two periods. Thus, in any period, a mass two of workers exists.

Workers are ex ante identical in this model. Thus, we suppress individual subscripts as we describe the preferences and production possibilities that characterize all workers.

Our model contains no information asymmetries or hidden actions. We do not contend that agents have no private information in these markets, but we do contend that young professionals begin their careers not knowing whether or not they possess the skills required for success in the role of partner. Our symmetric learning environment allows us to clearly highlight the connections between personnel policies and efficient ways to search for persons who can succeed as partners.

Because agents learn symmetrically in our model, we find it expedient to present our model as a planning problem. In section 6, we discuss market mechanisms that decentralize the solution to our planner's problem.

Our planner assigns workers to jobs and workloads. These workloads are the number of tasks that the planner assigns to each worker. Workers are risk neutral with the following utility function

$$U = m - c(n)$$

where m is expected income, n is the number of tasks performed, and $c(n)$ is the disutility of performing n tasks. We assume $c(0) = 0$, $c'(0) = 0$. Further, there exists a maximum workload, $\bar{n}$, s.t. $\lim_{n\to\bar{n}} c'(n) = \infty$, $c''(n) > 0 \quad \forall n \in [0, \bar{n}]$. All workers pay the same utility cost to complete any task.

Let θ denote worker ability, which is either high or low, i.e. $\theta \in \{0, x\}$, with $x > 0$. If a worker has high ability, the expected output generated by each professional task she completes is greater than the expected output generated by a low ability worker who performs the same task. At birth, the ability of workers is not known, but in each cohort, a constant fraction, π, is high ability, and the rest are low ability. All market participants know the distribution of ability.

Recall that there are two sectors. In both sectors, nature draws i.i.d. production shocks, ϵ, that are mean zero for all workers. In the outside sector, expected output is a linear function of worker effort, and the mapping between effort and output does not vary with worker experience or ability. Let w^o denote the marginal product of tasks performed in the outside sector. Outside sector output, y^o, is determined according to the following production function:

$$y^o = w^o n + \epsilon \tag{2.1}$$

Expected output in the professional sector is determined by worker ability, worker experience, and job assignment. Define y_s^j as the output of a worker assigned to professional job j given s periods of professional experience, where $j \in \{a, p\}$ for associate and partner, and $s \in \{0, 1\}$ for inexperienced and experienced. The production function for new associates is

$$y_0^a = (1 + \theta)n + \epsilon \tag{2.2}$$

The production function for experienced associates is

$$y_1^a = z^a(1 + \theta)n + \epsilon \tag{2.3}$$

Here, the parameter $z^a > 1$ captures the idea that associates who have experience are able to perform tasks better. Finally, the production function for experienced partners is

$$y_1^p = \begin{cases} z^p(1 + \theta)n + \epsilon & \text{if } \theta = x \\ -\infty & \text{if } \theta = 0 \end{cases} \tag{2.4}$$

The parameter z^p, where $z^p > z^a > 1$, captures the idea that partners perform tasks that more fully leverage professional skill. We assume that skill levels are functions of both experience and talent. Further, we assume that, if low ability workers of any experience level were to act as partners, the mismatch between their skills and their task assignments would create losses. To facilitate our exposition, we set the value of these losses to $-\infty$. Likewise, we assume that, regardless of their ability, workers with no experience would also make costly mistakes if they were to act as partners. So, we also set $y_0^p = -\infty$.

Our planner must allocate workers between the professional labor market and the outside sector. For now, we cap employment in the professional sector at $q < 1$ to capture the idea that only a fraction of highly-educated agents work in the professional sector. We also assume that the parameters of our production functions are such that the planner would always prefer to assign a new worker to the associate position than the outside sector, which implies that the cap, q, always binds.[9]

Later, we treat q as an endogenous variable that is determined by the costs of maintaining professional jobs and the productivities of positions in the professional sector.

We impose the following restrictions on production parameters:

---

[9]We assume constant returns to tasks performed in all sectors. So, without this assumption, it is possible that the planner would assign all new workers to the outside sector in each period. Thus, steady-state employment in the professional sector would be zero.

$$z^a < w^o < z^a(1 + \pi x) \tag{2.5}$$

The first inequality in (2.5) implies that an experienced associate who has low ability is more productive in the outside sector than in the professional sector. The second inequality implies that the expected productivity of an experienced associate with unknown ability is greater in the professional sector than in the outside sector. If this were not true, the planner's assignment decisions for experienced professionals with unknown ability would be trivial.

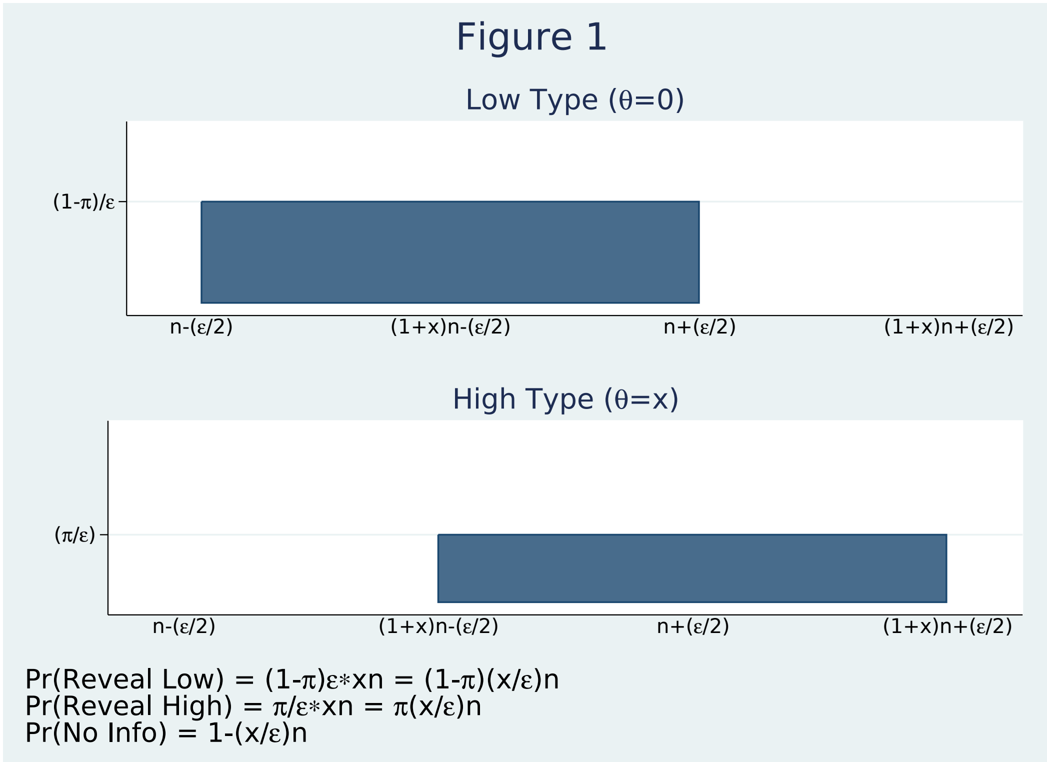
2.2. **Learning.** In our framework, the ability of each worker is drawn from a common distribution. During the first period, the market receives a signal about the ability of each new associate that either fully reveals her talent or provides no information about her talent. The likelihood that the market receives this fully revealing signal depends on the number of tasks an associate performs.

Define $\phi(n)$ as the probability that the market observes a revealing signal. We assume that $\phi(n)$ is an increasing and concave function of n, i.e. $\phi'(n) > 0$ and $\phi''(n) \leq 0$. Here, we provide one specific micro-foundation for this learning technology, but it is easy to construct others, and our key results would remain.[10]

Suppose the market observes the workload, n, and the resulting output for each professional worker, y_s^p. Then, given n, output provides a signal about worker ability. We follow Pries [2004] and assume that shocks to output are uniformly distributed, $\epsilon \sim U[\frac{-\varepsilon}{2}, \frac{\varepsilon}{2}]$. Figure 1 illustrates why this assumption generates all or nothing learning given our previous assumptions about y_s^p.

The two panels in the figure present two joint densities. Both densities describe output realizations for an associate who takes on a given workload, n. A given area under the density in the top panel equals the joint probability that a new associate both has low ability and produces output in a given interval. An area under the density in the bottom panel gives the corresponding joint probability of being high ability and producing output in a given range.

[10]See Bonatti and Horner [2017] and Bose and Lang [2017] for other frameworks where revealing signals arrive at rates determined by worker effort levels.



The regions of non-overlap between these joint densities contain signals that fully reveal the ability of new associates because only one ability type can produce the signals found in each of these regions. If a new associate produces less than $(1+x)n - \frac{\varepsilon}{2}$, the associate must be of low ability because a high ability associate would always produce at least this much. Further, if a new associate produces more than $n + \frac{\varepsilon}{2}$, the associate must be of high ability because a low ability associate would always produce this much or less.

Any signal in the region where these densities overlap provides no information about the ability of a new associate. For output values in this region, the joint density function in the bottom panel is $\frac{\pi}{\varepsilon}$ while the joint density in the top panel is $\frac{1-\pi}{\varepsilon}$. Thus, Bayes' rule implies that

$$Pr(\theta = x | y_0^a \in [(1+x)n - \frac{\varepsilon}{2},\ n + \frac{\varepsilon}{2}]) = \frac{\pi/\varepsilon}{\pi/\varepsilon + (1-\pi)/\varepsilon} = \pi$$

Given an output signal in the overlap region, the probability that a new associate is high ability is π, which is the prior probability that each new associate has high ability.

The length of this region of overlap, $[(1+x)n - \frac{\varepsilon}{2},\ n + \frac{\varepsilon}{2}]$, is $\varepsilon - xn$. Multiply this length by the density of the production shock, $\frac{1}{\varepsilon}$, to get, $1 - \frac{x}{\varepsilon}n$, which is the probability that the output signal reveals no information about associate ability.

Thus, $\phi(n) = \frac{x}{\varepsilon}n$ is the probability that an output signal reveals the type of a worker who takes on workload n. We assume that $\phi(\bar{n}) < 1$ to create an environment where it is not possible to achieve complete information about the ability of associates simply by working them "hard enough." Nonetheless, heavier workloads do create more information in this model, i.e. $\phi'(n) > 0 \; \forall n \in [0, \bar{n}]$. New associate effort, n, not only produces output but also reveals information about new associates.

Given our production environment, associates who reveal that they are low ability should always be re-assigned to the outside sector, and associates who reveal they are high ability should always be promoted to partner if they are retained. Optimal second period assignments for associates of uncertain ability are more subtle. Our analysis below highlights how z^p and z^a interact to determine both optimal workloads for new associates and whether or not experienced associates of uncertain ability face dismissal.

3. The Planner's Problem

Table 1 presents the five different types of workers in this economy. The rows of Table 1 correspond to the possible work histories. Since workers live two periods and can only work in one sector per period, a worker may have no experience in either sector, one period of experience in the outside sector, or one period of experience in the professional sector.

The columns of Table 1 correspond to the three possible information states concerning the ability of a worker. The market may know that the worker has low ability ($\theta = 0$) or high ability ($\theta = x$), and the market may be uncertain, i.e. retain the common prior that applies to all new workers.

The combination of the three experience types and three information sets yield the nine cells in Table 1. The first two columns of the first two rows are marked "n.a." since by assumption the ability of workers with no professional experience is uncertain.

The planner must choose a job assignment and workload for each of the five types that remain. Many of these choices are trivial. The optimal job assignment for three of these types is immediate, and given all but one potential job assignment, optimal workloads are solutions to simple static optimization problems.

Table 1
Optimal Assignment for Worker Types

		Ability		
		$\theta = 0$	$\theta = x$	$Pr(\theta = x) = \pi$
History	New	n.a.	n.a.	Associate / outside
	Experienced Outside	n.a.	n.a.	Outside
	Experienced Professional	Outside	*Partner*	?

Notes: The rows delineate three types of workers: new, experienced in the outside sector, and experienced in the professional sector. The columns spell out the three possible information states about worker ability.

Let us start by considering the three trivial assignment decisions. First, recall that, $y_0^p = -\infty$. Thus, the planner never assigns new workers to the partner position. Further, since $q < 1$, the planner cannot assign all new workers to the associate position. Thus, he assigns some fraction of them to work as associates in the professional sector and assigns the rest to the outside sector.

Second, the planner assigns workers with experience in the outside sector to remain in the outside sector. Clearly the planner would never assign them to work as partners, since he cannot rule out the possibility that they have low ability, $x = 0$. But, the planner should also not assign them to work as associates. New workers have the same expected productivity in the associate position, and they may turn out to be partner material.

Third, the planner assigns workers revealed to be low ability to the outside sector. These workers are more productive in the outside sector given $z^a < w^0$. See equation (2.5).

This leaves the two assignment decisions. The planner must assign experienced professionals whose ability is either uncertain or known to be high. Below, we show that after at most one period, the planner would assign all experienced workers known to be high ability to the partner job.[11] This leaves the question of where to assign experienced workers with uncertain ability: should the planner leave them in the professional sector or assign them to the outside sector? If the latter holds, we say that the planner follows an up-or-out rule.

The planner must also assign workloads to each of the five worker types. The planner assigns the same workload to outside workers whether they are experienced or inexperienced. Output in the outside sector produces no information about worker talent, and once workers enter the outside sector, they never leave. So, the optimal workload for all outside workers is the one that maximizes current period surplus. Further, the planner assigns workloads to experienced professionals that maximize the surplus they create in the last period of life.

[11]The qualifier, "after the first period," addresses the possibility that the initial endowment of high-ability, experienced professionals is greater than q. Because q limits the number of new associates in any period, the planner always has less than q professionals with known high-ability after the first period.

In contrast, the optimal workload for new associates is not the solution to a simple static problem. Among new associates, heavier workloads generate more output and more information about ability. Because this information guides assignment decisions in the next period, the optimal workload for new associates reflects the fact that effort today affects assignments and workloads in the future.

3.1. **Elements of the Planner's Problem.** Given these results, we can state the planner's objective function. Recall that we restrict our attention to production parameters such that the planner wants to fill all q positions in the professional sector. Thus, he will employ $2-q$ workers in the outside sector, each of whom will produce a surplus of

$$v^o \equiv \max_n \;\; w^o n - c(n)$$

Let ρ^x denote the mass of experienced workers known to be high ability. The planner chooses the fraction α^x of such workers to retain and produce a surplus of

$$v_1^p \equiv \max_n \;\; z^p (1+\theta) n - c(n)$$

Let ρ^u denote the mass of experienced workers with unknown ability. The planner chooses the fraction α^u of such workers to retain and produce a surplus of

$$v_1^a \equiv \max_n \;\; z^a (1+\pi\theta) n - c(n)$$

Finally, the remaining $q - \alpha^u\rho^u - \alpha^x\rho^x$ workers are inexperienced workers who the planner assigns to work as new associates. The surplus they produce when young is

$$v_0^a \equiv (1+\pi\theta) n - c(n)$$

where n is their workload. Hence, the flow surplus the planner creates given his choices of α^u, α^x, and n is given by

$$s(\alpha^u, \alpha^x, n) \equiv (2-q) v^o + q v_0^a(n) + \alpha^x\rho^x \left(v_1^p - v_0^a(n)\right) + \alpha^u\rho^u \left(v_1^a - v_0^a(n)\right)$$

The planner's objective is to choose workloads and job assignments that maximize the present discounted expected value of the infinite stream of per-period surplus generated in this economy. If we assume that the planner discounts the future using $\beta < 1$, we can write the planner's problem using the following recursive formulation

$$V(\rho^u, \rho^x) = \max_{\alpha^x, \alpha^u, n} s(\alpha^u, \alpha^x, n) + \beta V\left(\rho^u_{+1}, \rho^x_{+1}\right) \tag{3.1}$$

where

$$\begin{aligned} \rho^u_{+1} &= (q - \rho^u\alpha^u - \rho^x\alpha^x)(1-\phi(n)) \\ \rho^x_{+1} &= (q - \rho^u\alpha^u - \rho^x\alpha^x)\pi\phi(n) \end{aligned}$$

as well as the constraint

$$q - \rho^u\alpha^u - \rho^x\alpha^x \geq 0 \tag{3.2}$$

This last constraint states that the planner cannot retain more experienced professional workers than there are slots in the professional sector. Note there can be at most q new associates in any period, and the planner always learns that a fraction of these new associates have low ability. Thus, in any period after the initial period, $\rho^u + \rho^x < q$, and constraint (3.2) cannot bind. In the analysis that follows, we discuss the planner's optimal choices in this region of the state space.

Let $\{\hat{\alpha}^u, \hat{\alpha}^x, \hat{n}\}$ denote the solution to planner's problem. If the planner begins with positive stocks of uncertain and high-ability professionals, $\rho^u > 0$ and $\rho^x > 0$ that satisfy constraint (3.2), the planner faces a trade off. These workers are more productive than new associates. The uncertain types are more productive because they have more professional experience. The high ability types are not only more skilled but also able to work in partner positions that exploit their skills. However, for each experienced professional that the planner retains in the professional sector today, he will have one less experienced professional next period. Further, the planner's effort choice for new associates, n, interacts with these retention decisions because the probability that the planner observes the actual ability of a new associate is $\phi(n)$.

3.2. **Promotion to Partner.** To begin, we establish that $\hat{\alpha}^x = 1$, i.e. the planner retains all experienced professionals with known high ability. Given our assumptions about z^p, the planner not only retains these professionals but also promotes them all to partner.

As a first step, note that there are no productivity spillovers among workers in this model. We therefore expect $V(\rho^u, \rho^x)$ to be linear. We confirm that this is the case and use linearity to establish that the planner always chooses $\hat{\alpha}^x = 1$.

Claim 1: $V(\rho^u, \rho^x) = K_1 + K_2\rho^u + K_3\rho^x$ for all (ρ^u, ρ^x) s.t. $0 \leq \rho^u + \rho^x < q$.

See Appendix A for the proof. The linearity of $V(\rho^u, \rho^x)$ reflects that fact that the value created by experienced professionals of high ability is not influenced by the stock of experienced professionals of uncertain ability, and vice versa. Further, the value of an addition to either stock of experienced professionals does not depend on the current level of either stock. These observations form the basis of our second claim.

Claim 2: $\hat{\alpha}^x = 1$ for any pair (ρ^u, ρ^x) s.t. $0 \leq \rho^u + \rho^x < q$.

Since the value generated by an experienced professional of high ability is independent of the current stock variables, the planner gains nothing from trying to smooth the stock of experienced, high-ability professionals over time. This implies that he never replaces a high-ability, experienced professional with a new associate. If the planner were to make such a replacement, there would be a probability, $\pi\phi(\hat{n}) < 1$, that the new associate would be revealed to have high-ability at the end of the period, and even in this case, she would be no more productive next period than a high-ability, experienced professional would be this period.

4. The Link Between Workloads and Up-or-Out

We have now determined optimal rules for eight of the ten choices the planner makes each period. There are two choices that remain: the workload for new associates, n, and the job assignment for experienced associates of uncertain ability, α^u. In this section, we demonstrate how these choices are related.

Substitute the expression for $V(\rho^u, \rho^x)$ in Claim 1 into equation (3.1). Then, take the derivative with respect to n to get the first order condition that defines the optimal workload for new associates, $\hat{n}$:

$$c'(\hat{n}) = (1 + \pi x) + \beta\phi'(\hat{n})(\pi K_3 - K_2)$$

This equation highlights an important property of optimal workloads for new associates. The marginal cost of new associate effort must equal the sum of two marginal returns. The first, $(1 + \pi x)$, is the expected marginal product of new associate effort. The second, $\beta\phi'(n)(\pi K_3 - K_2)$, is the marginal information return from worker effort.

For the planner, K_2 represents the value created by replacing a new associate with an experienced professional who possesses uncertain ability, while K_3 represents the value created by replacing a new associate with an experienced professional who has known high ability. When the planner marginally increases a new associate's workload, n, the probability that the associate's output signal reveals her ability increases by $\phi'(n)$. If the signal is revealing, there is a probability π that the signal will reveal high ability, and the planner's value function will increase by K_3 instead of K_2. Thus, $\phi'(n)(\pi K_3 - K_2)$ is the marginal information rent generated next period by new associate work this period.

We show in Appendix A that πK_3 is always greater than K_2. Thus, the information rents created by new associate effort are always positive. Given this result, we prove the following proposition:

Proposition 1. *The optimal workload for new associates, $\hat{n}$, exceeds the static optimum implied by the expected per-period output of new associates.*

New associates take on workloads that are greater than those that would maximize the current surplus generated by their positions in order to produce information that improves professional job assignments in the future.

A similar trade off between current output and future information shapes the planner's decision concerning the retention of experienced associates of uncertain ability. The first order condition that defines $\hat{\alpha}^u$ is

$$\hat{\alpha}^u = \begin{cases} 1 & \text{if } v_1^a - v_0^a(\hat{n}) - \beta\left[(1-\phi(\hat{n}))\,K_2 + \pi\phi(\hat{n})K_3\right] \geq 0 \\ 0 & \text{if } v_1^a - v_0^a(\hat{n}) - \beta\left[(1-\phi(\hat{n}))\,K_2 + \pi\phi(\hat{n})K_3\right] < 0 \end{cases}$$

In Appendix A, we solve for K_2 and K_3 and use these solutions to express this condition in a more informative way.

Proposition 2. $\hat{\alpha}^u = 0$ *if* $v_1^a - v_0^a(\hat{n}) < \beta\pi\phi(\hat{n})(v_1^p - v_1^a)$

The left-hand side of the inequality in Proposition 2 is the current surplus cost of replacing an experienced associate of unknown ability with a new associate. The right-hand side gives the expected future returns that these replacements create. The probability that new associates become partners next period is $\pi\phi(\hat{n})$, and the additional surplus generated by partners is $v_1^p - v_1^a$.

Thus, the planner's decision concerning $\hat{\alpha}^u$ reflects a trade off between the current value of worker experience in the associate position and the discounted expected value of identifying more high-ability workers in the future and promoting them to the partner position. Up-or-out rules are optimal in environments where the increase in expected future surplus associated with increasing the number of future partners outweighs the loss in current surplus that comes from replacing experienced associates with new ones.

We are interested in how the relative productivities of different types of workers in different roles within organizations shape optimal personnel policies. Thus, we want to understand how z^p and z^a shape the value of identifying candidates for promotion to partner and the surplus cost associated with replacing experienced professionals. Our key comparative static results spell out how both parameters affect $\hat{n}$ and $\hat{\alpha}^u$.

Proposition 3. *The optimal workload for new associates, $\hat{n}$, is increasing in z^p and weakly decreasing in z^a.*

The probability that a new associate's output signal reveals her true ability increases with $\hat{n}$. Thus, if new associates work more this period, the planner will be able to identify and promote more partners next period. For parameter values such that $\hat{\alpha}^u = 1$, the additional surplus generated by these promotions increases with z^p and decreases with z^a. Therefore, optimal effort, $\hat{n}$, increases with z^p and decreases with z^a. If $\hat{\alpha}^u = 0$, z^a does not enter these surplus calculations because no one works as an experienced associate. However, $\hat{n}$ still increases with z^p.

We argue that many professional service firms employ both heavy workloads for new associates and up-or-out promotion rules as tools that facilitate their search for talented partners. Thus, the effects of z^p and z^a on firm decisions concerning up-or-out should be similar to their effects on workloads for new associates. Our second comparative static result confirms this:

Proposition 4. *$\hat{\alpha}^u$ is weakly decreasing in z^p and weakly increasing in z^a.*

Up-or-out is optimal when the option value associated with a new associate exceeds the productivity gains that come from associate experience. The returns from finding talented professionals and promoting them to partner are increasing in z^p, and z^a determines the gains from associate experience.

Taken together, Propositions 3 and 4 imply that, given certain parameters, firms adopt both up-or-out promotion rules and heavy workloads for new associates. Given any initial pair (z^a, z^p), $\hat{n}$ increases and an up-or-out rule is either retained or adopted as we increase z^p. Further, $\hat{n}$ decreases and an up-or-out rule is either retained or abandoned as we increase z^a.

In settings where both z^a and z^p are relatively large, associates may work long hours without facing up-or-out. Still, in environments where z^p is large enough relative to z^a, the planner chooses heavy workloads and up-or-out retention rules. These policies make no sense as static allocation rules. Yet, both policies are optimal because the planner is not just producing surplus for the current period. He is also conducting a search for talent, and the results of this search impact future surplus.

4.1. **Related Empirical Evidence from Professional Labor Markets.** Several authors point out that strict adherence to up-or-out rules became less common in law firms during the 1980s and 1990s.[12] Gorman [1999] argues that, as the market for legal services grew, private law firms began doing relatively more work that required special expertise and experience. Mergers, changes in information

[12]See Gilson and Mnookin [1985], Galanter and Palay [1991], Gorman [1999], and Galanter and Henderson [2008].

technology, and increases in the demand for specialized legal services allowed larger law firms to create a limited number of new non-partner roles for experienced specialists. Among lawyers who were well-suited to these new roles, z^a increased, and it became optimal to retain these lawyers even if they were not well-suited to the partner role.

As the market for professional services has grown in recent decades, the relative demand for specialists has produced similar movements away from strict up-or-out policies in public accounting, especially in the largest accounting firms. Press reports and studies by the American Institute of Certified Public Accountants (AICPA) indicate that, in recent decades, a significant fraction of public accounting firms have created terminal senior roles that are not partnerships.[13]

This development in public accounting is of particular interest to us because press reports and AICPA studies also report a parallel trend toward less demanding work schedules in public accounting. These reports do not contain any information on trends in total hours worked among public accountants at different points in their careers. However, these sources do document trends in the adoption of a number of practices that seek to reduce work related stress. In recent decades, many accounting firms have increased the number of paid holidays their professionals receive, increased the use of job sharing arrangements, increased the use of flexible scheduling, and given professionals expanded opportunities to work from home and avoid commuting time.[14]

Based on extensive reading of the trade literatures, we conclude that, in most elite law and public accounting firms, the majority of young professionals still begin their careers expecting that they will either move up to partner or out to another employer. However, as mergers have created larger firms in these industries, some firms have created a limited number of senior roles for experienced professionals with special expertise, and in public accounting, this development has been accompanied by the adoption of personnel policies that appear to make the workloads of young accountants less burdensome.

We have not identified any data sets that would allow us to directly investigate the extent to which our model provides the correct explanation for these trends, and such data may not exist. Propositions 3 and 4 above deal with changes in z^a or z^p holding all else constant, but the waves of mergers and changes in information technology that affected law and accounting firms in recent decades likely affected the productivity of new associates, experienced specialists, and partners simultaneously.

The predictions of our model may map more cleanly into long-standing differences between the labor market for professional services and other markets for well-educated workers. While there is some evidence that careers in public accounting are becoming less stressful, the trade literatures on professional labor markets still indicate that most new entrants in professional service firms continue to compete in up-or-out promotion contests while working more arduous schedules than their

[13]Press accounts concerning changes in the use of up-or-out rules in public accounting firms echo Gorman's claims about the rising value of specialists in law. See **New York Times**, May 17, 1990. See Almer [2004] for more recent data from the AICPA. Directorships are common positions that involve permanent senior roles without partnership status. Alternative partnerships also exist. These positions do not involve the same equity status as regular partners.

[14]See Table 23 in Almer [2004]. See also Greenhouse [2011] and Lewison [2006].

peers in other industries.[15] Our model suggests that this is because an entry level professional in a major manufacturing or service firm is not really auditioning to be CEO or even a division president. While the largest professional service firms have thousands of partners, the largest traditional corporations have only handfuls of people in their highest leadership roles. Further, while professional service firms offer a limited number of senior roles for non-partners, large corporations contain many productive roles for experienced professionals who do not reach the absolute highest levels of management.

Viewed through the lens of our model, an experienced professional in a traditional firm enjoys a large effective value of z^a because the firm can assign her to the one role, among many potential roles, that suits her best. This implies that, in traditional firms, there is no urgency to determine whether or not a given young professional is well-suited to one specific senior role. In contrast, a new entrant in a professional services firm is auditioning to be a partner in the firm and expects to leave the firm if the audition for that one role does not go well. In this setting, it makes sense to expedite the process of determining whether or not this new entrant is actually partner material, and it makes sense for her to choose a different career path if she learns that she is not well-suited to the partner role.

4.2. **Stay or Go.** A large literature on screening rules points out that, in models with one position, the option value associated with bringing in a new worker determines the stringency of the rule that governs the retention decision for incumbent workers with uncertain ability. Thus far, we have assumed $z^p > z^a$, but if $z^p = z^a$, the professional sector in our model effectively contains only one position. Once again, the planner will dismiss all experienced professionals who are known to possess low-ability, $\theta = 0$, and retain all experienced professional with known high-ability, $\theta = x$. Given some parameters, $\hat{\alpha}^u = 1$ and given others, $\hat{\alpha}^u = 0$, but in both settings the planner is only making a retention decision. There is no promotion decision.

In this setting, some may describe the more stringent screening rule, $\hat{\alpha}^u = 0$, as an up-or-out policy, but this is not how we employ the term. When $z^a = z^p$, one can interpret $\hat{\alpha}^u = 0$ as a decision to search for stars, but our goal is to examine the links between personnel policies and the relative productivities of different positions in an organization.

Partner is not simply a title that firms give to experienced associates who produce high-quality work. Partners in professional service firms also perform the business development and client management activities that allow their firms to survive and grow. Our assumptions concerning z^p, z^a, and the losses that would occur if low ability types were to occupy the partner position are all attempts to model the process of searching for persons who can succeed in the demanding role of partner.

5. Learning About Rates of Learning

In our model, new associates take on heavy workloads in order to produce information about their suitability for the partner role. Yet, both the academic and trade literatures on professional labor markets assert that young professionals take

[15] For example, Bertrand et al. [2010] report that recent MBA graduates who enter investment banking and consulting appear to longer hours than their peers who enter corporate jobs.

on these heavy workloads for an additional reason. New associates learn by doing, and their heavy workloads allow them to acquire valuable skills faster.[16]

Therefore, we now consider human capital accumulation. But, just as we assume that workers do not begin their careers with a complete understanding of their own skill levels, we also assume that workers do not begin their careers with certainty about their capacity to acquire new skills. This implies that they only learn ex post whether or not their work experiences have produced the skills required to earn promotions or raises. In this section, we flesh out the many parallels between learning about worker skill and learning about the ability of workers to acquire skills. We show that both frameworks allow us to model professional service firms as organizations where new associates audition for partner positions. Appendix B presents a fully-specified model and analyses it. Here, we summarize key features and results.

In our model above, new associates possess different ability levels, and three information states are possible for an associate who has just completed her first period of professional work. The planner may know that she has high ability. The planner may remain uncertain about her ability, or the planner may know she has low ability.

In our learning by doing model, new associates begin their careers at a common skill level but learn by doing at different rates that are unknown to them or the planner.[17] To facilitate comparisons with our screening model, again assume that three information states are possible for an associate who has just completed her first period of professional work. The planner may know that she has acquired the skills needed to function well as a partner. The planner may know that she has acquired the skills required to function effectively as an experienced associate but not as a partner, or the planner may know she has obtained few skills and belongs in the outside sector.

In our screening model, with probability $\pi\phi(n)$, associates who perform workload n learn that they have high ability. With probability $1-\phi(n)$, these associates learn nothing and maintain their prior beliefs about their abilities. With probability $(1-\pi)\phi(n)$, they revise their beliefs downward because they discover that they possess low-ability.

In our learning by doing model, no one observes, ex ante, the different rates at which new associates learn by doing. Yet, ex post, all market participants do observe the skill level that each experienced associate has obtained. Since skills grow at uncertain rates, the workload, n, that the planner assigns to new associates determines the distribution of realized skill levels among experienced associates, and $\phi_h(n)$, $\phi_m(n)$, and $\phi_l(n) = [1-\phi_m(n)-\phi_h(n)]$ are the probabilities that a new associate who performs n tasks achieves a high, medium, or low skill level, respectively.

In our screening model, our key assumption about the information technology is that the probability of a revealing signal, $\phi(n)$, is an increasing and concave function of n. This implies that the probability of identifying a high-ability worker, $\pi\phi(n)$,

[16]See Rosen [1972] for an early model of the market for training opportunities. See Wilkins and Gulati [1998] for a discussing of how partners train law associates. See Batchelor [April 20, 2011] for a discussion of training in management consulting.

[17]Demougin and Siow [1994] present another model of a training and screening where new professionals train for management positions, but ex post, only some acquire the skills required to be effective managers.

is increasing and concave, while the probability of remaining uncertain about the ability of an experienced professional is decreasing in n. In our learning by doing model, the analogous conditions are that $\phi_h(n)$ is increasing and concave and $\phi_m(n)$ is decreasing in n.

Given these conditions, our learning by doing model is essentially isomorphic to our screening model, and we are able to establish results that parallel all of our propositions above.[18] Our model is a model of auditions. Young professionals enter associate positions and work long hours as part of a process that reveals whether or not they either have the skills or can acquire the skills they need to perform well as partners.

Appendix B demonstrates that all the results from our screening model carry over to our model of learning by doing if $\phi_h(n)$ is increasing and concave and $\phi_m(n)$ is weakly decreasing. Yet, because young associates acquire more skills when they put in more effort, additional effort could, in principle, increase both $\phi_m(n)$ and $\phi_h(n)$.[19] In these cases, results that parallel Propositions 1 through 4 still hold as long as the ratio $\phi_h(n)/\phi_m(n)$ increases with n sufficiently fast. If this ratio does not rise fast enough with n, the comparative static results in Proposition 3 may flip, i.e. n may decrease with z^p or increase with z^a. Further, the retention rule for professionals who achieve the medium skill level may exhibit non-monotonicities in z^p and z^a.

In sum, Appendix B establishes results that parallel those from our screening model under the additional restriction that the ratio $\phi_h(n)/\phi_m(n)$ must increase with n at rate that is sufficiently fast. Given this restriction, the relative productivities of partners and experienced associates shape the trade offs that determine promotion rules and workloads just as they do in our screening model above.

6. Endogenous Sector Size and Decentralization

So far, we have described solutions to a planner's problem given the constraint that professional employment cannot exceed a binding cap of q. Here, we explain why our results hold in a more general setting where the planner also chooses the optimal size of the professional sector. Then, we explain why an equilibrium of a decentralized economy implements the solution to this more general planner's problem. Appendix C proves these results.

6.1. Endogenous Sector Size.

We do not model interactions among professionals who work in the same firm. Therefore, our model does not address optimal firm size or the organization of work in firms. We only characterize the optimal size of the professional sector.

[18] Our learning by doing model is not identical to our screening model. In the screening model, the probability that the planner learns that a new associate has high ability is linearly related to the probability that the planner receives no information about the associate, but in our learning by doing model, there is no linear relationship between the corresponding probabilities $\phi_h(n)$ and $\phi_m(n)$. Moreover, in our screening model, the probability that the planner identifies an associate with low ability is increasing in n while the corresponding probability in our learning by doing model, $\phi_l(n)$, is increasing in n. Nonetheless, these differences do not matter for the arguments we use to prove Propositions 1 through 4 above.

[19] How $\phi_m(n)$ varies with n depends jointly on the distribution of learning efficiencies among new associates, the skill requirements for different positions, and the cost function for worker effort. See Appendix B.

Professional workers require office space, support staff, and other resources that facilitate their capacity to interact efficiently with clients. We assume that the supply curve for effective support services to the professional sector is upward sloping. As the professional sector grows, it must employ support staff who have better outside options and occupy office space that has more valuable alternative uses.

Suppose the per-period cost of supporting the q-th professional position is given by $\kappa(q)$, where $\kappa(\cdot)$ is an increasing continuous function such that $\lim_{q\to 0}\kappa(q) = 0$ and $\lim_{q\to 1}\kappa(q) = \infty$.[20] Given our previous assumptions, these restrictions ensure that the planner assigns some but not all of the workers in a cohort to begin their careers in the professional sector.

The planner faces a problem similar to the one described in equation (3.1), but here, he must also determine $\hat{q}$, the optimal number of professional workers. Let α^x denote the fraction of experienced professionals with known high ability that the planner retains. Appendix C shows that, with the possible exception of the initial period, the planner always retains these workers and promotes them to partner, just as in the fixed q problem.[21] However, explicit notation for this decision allows us to more clearly formulate our new planner's problem:

$$V\left(\rho^u, \rho^x\right) = \max_{n,\alpha^u,\alpha^x,q} \quad s(n, \alpha^u, \alpha^x) - \int_0^q \kappa(y)dy + \beta V\left(\rho_+^u,\ \rho_+^x\right) \tag{6.1}$$

$$\begin{aligned} \text{s.t.} \quad & (1) \quad \rho_+^u = \left[q - \alpha^u\rho^u - \alpha^x\rho^x\right]\left[1 - \phi(n)\right] \\ & (2) \quad \rho_+^x = \left[q - \alpha^u\rho^u - \alpha^x\rho^x\right]\pi\phi(n) \\ & (3) \quad q - \alpha^u\rho^u - \alpha^x\rho^x \geq 0 \end{aligned} \tag{6.2}$$

As before, constraints (1) and (2) are the laws of motion for the state variables. Constraint (3) states explicitly that the number of professional positions the planner creates must weakly exceed the number of experienced professionals he retains.

Appendix C shows that, with the possible exception of the initial period, the planner chooses an optimal sector size, $\hat{q}$, that is constant over time. Appendix C also shows that, if the initial stocks of experienced professionals are so large that the planner chooses a sector size greater than or equal to $\hat{q}$ in the initial period, constraint (3) will bind, and the planner will assign all new workers to the outside sector in the initial period. Taken together, these results imply that there can be at most $\hat{q}$ new associates in any period. Hence, beyond the initial period, the stock of experienced workers can be at most $\hat{q}$, and some will be known to possess low ability. This implies $\rho^u + \rho^x < \hat{q}$, so constraint (3) never binds after the initial period.

Since the productivity of a worker in a given position is not influenced by how other workers are allocated to positions, optimal workloads, retention rules, and

[20] $\kappa(q)$ is the cost of supporting a professional position. We could also include support costs that vary with workloads n as long as total support costs for a given professional are separable in q and n. In this case, we can think of variable support costs as an additional component of $c(n)$ for workers in the professional sector.

[21] If the initial stock of skilled professionals, ρ_0^x, is sufficiently large, the planner may not retain all of them in the professional sector. However, we show in Appendix C that, after the initial period, the planner promotes all skilled professionals to partner.

promotion policies are the same whether there are five professional positions or five million. For this reason, the planner chooses $(\hat{n}, \hat{\alpha}^u, \hat{\alpha}^x)$ using the same rules he employed in the fixed q problem, and all of the comparative statics results that we derived in section 4 continue to hold.

However, the fact that the mapping between production parameters, (z^a, z^p), and optimal personnel policies is independent of sector size does not imply that changes in (z^a, z^p) have no effect on the planner's sector size choice, $\hat{q}$. Appendix C proves the following proposition:

Proposition 5. *The optimal sector size, $\hat{q}$, is increasing in z^p and weakly increasing in z^a.*

The relative values of z^a and z^p influence personnel policies, but the levels of z^a and z^p also determine the value of professional sector work. When professional work is more valuable, the planner optimally chooses a higher $\hat{q}$ and allocates fewer workers to the outside sector. Note that $\hat{q}$ only increases weakly with z^a. In the region of the parameter space where up-or-out is optimal, no one works as an experienced associate, and z^a has no impact on professional sector surplus.

6.2. **Decentralization.** Since workers have no private information about their abilities or their actions, many different market mechanisms could decentralize the solution to our planner's problem. We show that one particular mechanism does. We do not explore the details of the proofs here. Rather, we develop insights concerning how and why this mechanism would work, and we use these insights to offer specific interpretations of both up-or-out policies and the heavy workloads new associates bear while competing in up-or-out promotion contests.

Assume that all workers choose whether to work in the outside sector at a fixed wage, w^o, or work in the professional sector as independent contractors.[22] Whether these independent professionals choose to work in the associate or partner role, we assume that each must hire support services at a cost determined by the equilibrium size of the professional sector.[23] Appendix C demonstrates that the market equilibrium in this independent contractor scenario implements the planner's solution to the problem described by equation (6.1). Here, we sketch the argument in two steps. First, we consider how independent contractors choose workloads and jobs when they believe the sector size is going to equal the planner's choice $\hat{q}$. Then, we argue that if all contractors follow these rules, rational entry decisions will generate a professional sector of size $\hat{q}$.

As we note above, the optimal sector size, $\hat{q}$, is the same in all periods, with the possible exception of the initial period. Thus, in our decentralization, workers make their hours and job choices assuming that the size of the professional sector is fixed over time. Define $\upsilon(n, q)$ as the expected lifetime utility of a new associate who chooses workload n given a professional sector size q.

[22]We treat workers as independent contractors to facilitate exposition. The same results would hold in a competitive labor market where identical professional service firms posted a menu of employment contracts that specified optimal workloads, termination rules, and wages equal to expected marginal products for all possible combinations of worker types and positions.

[23]Competition for staff workers implies that all professionals must pay their staff the outside option of the marginal staff worker.

$$(6.3) \qquad \begin{aligned} \upsilon(n,q) &= v_0^a(n) - \kappa(q) + \beta\pi\phi(n)[v_1^p - \kappa(q)] + \\ &\quad \beta(1-\pi)\phi(n)v^o + \beta(1-\phi(n))\max[v^o, v_1^a - \kappa(q)] \end{aligned}$$

The first two terms are the expected utility of working as an associate. The next term is the worker's expected discounted utility from learning that she has high ability and is therefore suitable for the partner role. The fourth term is the probability that she learns that she has low ability multiplied by the discounted value of working in the outside sector in the second period. The last term is the expected discounted value of being uncertain about her ability at the end of the first period.

Now, consider $\upsilon(n,\hat{q})$, and ask what choice of n is optimal for a new associate who believes that the size of the professional sector is the planner's choice $\hat{q}$. Since associates who become partners receive the surplus produced by the partner position, they internalize the information rents produced by their first period workload. So, it is no surprise that Appendix C shows that $\hat{n}$ maximizes $\upsilon(n,\hat{q})$.

Next, consider the last term of $\upsilon(n,q)$ in equation (6.3). Appendix C also demonstrates that the planner imposes an up-or-out rule, $\hat{\alpha}^u = 0$, if and only if experienced independent professionals with uncertain ability choose the outside sector when facing a professional sector size of $\hat{q}$, i.e. $v^o > v_1^a - \kappa(\hat{q})$. To gain intuition for this result, recall that, without regard to sector size, the planner chooses $\hat{\alpha}^u = 0$ in all environments where new associates are more valuable than experienced associates. Further, the planner's choice of $\hat{q}$ and therefore $\kappa(\hat{q})$ equates the value of a new associate and an outside worker. If outside workers are just as valuable as new associates and new associates are more valuable than experienced associates, we expect an experienced professional who is not partner material to find outside work more attractive than the experienced associate position.

We have shown that our independent contractors choose the planner's workloads and follow the planner's job assignment rules if they face support costs $\kappa(\hat{q})$. To complete the decentralization argument, we now show that, if all workers correctly assume that independent professionals are following the planner's personnel policies, free entry generates a professional sector of size $\hat{q}$.

Assume that all contractors follow the planner's personnel policies and that q is a time-invariant equilibrium sector size. These assumptions imply that $\upsilon(\hat{n},q) = v^o(1+\beta)$. Otherwise, all new workers would have a strict preference for either the professional or outside sector, and this cannot be true in equilibrium. Given our assumptions concerning $\kappa(\cdot)$, $\upsilon(\hat{n},q)$ declines continuously and monotonically with q, and there can be at most one q such that $\upsilon(\hat{n},q) = v^o(1+\beta)$. Appendix C proves that $\upsilon(\hat{n},\hat{q}) = v^o(1+\beta)$, which we expect since the planner chooses $\hat{q}$ to equate the value of associate positions and outside work.

Since the planner's solution, $(\hat{n},\hat{\alpha}^u,\hat{q})$, is an equilibrium of our independent contractor equilibrium, the comparative statics results we present in Propositions 1 through 5 hold in our independent contractor economy as well. However, several features of this equilibrium point to new ways to interpret these results.

The indifference condition $\upsilon(\hat{n},\hat{q}) = v^o(1+\beta)$ always holds in our contractor equilibrium, and it implies that the market price for support staff, $\kappa(\hat{q})$, always equals the maximum amount that new associates are willing to pay to work in the professional sector. Given this observation, consider our comparative static results concerning how changes in z^a and z^p affect the use of up-or-out. In an up-or-out equilibrium of our contractor economy, new associates are willing to pay more to work in the professional sector than experienced professionals who remain uncertain about their talent. Thus, new associate demands for support staff price these experienced professionals out of the professional sector. From this starting point, if we increase z^a enough to make these experienced professionals willing to pay as much or more than new associates are willing to pay for support staff, we can get them to stay and work as experienced associates.

On the other hand, if we start at an equilibrium without up-or-out, and we increase z^p, we increase the option value inherent in new associate positions while holding the value of working as an experienced associate constant. This increase in option value increases the amount that new associates are willing to pay in support costs to enter the professional sector, and since the demands of new associates determine the market price of support staff, we can always price experienced, non-partner professionals out of the market by making z^p sufficiently large.

Our model treats new associate work as an audition for the partner role, and in our contractor economy, the price of support staff clears this market for auditions. However, $\kappa(\hat{q})$ is not the only audition cost that new associates pay. New associates also choose to take on heavy workloads, and as a result, they enjoy lower first-period utility than outside workers.[24]

This insight may help explain why young professionals often report low job satisfaction, and in particular, why they report that they would be willing to accept lower earnings in exchange for less demanding work schedules.[25] Suppose that young professionals who respond to such surveys are reporting that they are willing to accept lower earnings in exchange for less demanding workloads, holding all else constant, including their future prospects for promotion. If so, new associates in our model would express a willingness to exchange current salary for reduced workloads. The problem is that there is no way to make such an exchange while holding all else constant. If new associates did perform fewer tasks, the market would learn less about them, and they would be less likely to become partners.

Landers et al. [1996] stress that young lawyers are less willing to report that they want to reduce their hours if they are informed that their peers do not want to reduce their hours. This pattern is clearly consistent with the coordination role that norms play in rat-race models. However, once an enumerator has informed a young professional that her peers chose not to reduce their hours, many mechanisms that link work effort and promotion probabilities may become more salient for the respondent.

Some scholars have argued that up-or-out policies are puzzling because surely some experienced associates who are not well suited to the partner role are nonetheless competent professionals. Why would firms refuse to negotiate a retention package for these associates?[26] Our interpretation of associate jobs as auditions for partnerships provides an answer. In our framework, young professionals in up-or-out

[24]This follows directly from $v(\hat{n}) = v^0(1 + \beta)$, and in part, reflects the fact that $\hat{n}$ is greater than the workload that maximizes static surplus.

[25]See Landers et al. [1996] for survey responses from young lawyers.

[26]See Kahn and Huberman [1988]. Gilson and Mnookin [1989] discuss this puzzle in the context of experienced legal associates. Batchelor [April 20, 2011] discusses how leading consulting firms often place those who fail to make partner in prestigious jobs in large companies that are clients of the firm.

markets pay a utility cost to acquire information about whether or not they are well-suited to lucrative partner positions. Thus, when a young professional learns that she is not partner material, she is not willing to pay this cost any longer. Further, although she would be more productive, in expectation, than the new associate who replaces her, her firm cannot profitably make a retention offer that she would accept because it can no longer offer her the opportunity to learn about her fitness for the partner role.

7. Empirical Patterns Concerning Hours and Promotions

In this section, we explore data that describe career outcomes among young lawyers. We focus on differences in hours worked, billing rates, and total compensation among lawyers who occupy different positions during the first twelve years of their careers, and we highlight one pattern that appears in two different data sources.

On average, lawyers who begin their careers as associates in private law firms and do not make partner typically reduce their work hours significantly when they leave the partnership track. These hours reductions occur among the small number who take positions in private law firms that are not on the partnership track and among the much larger number who accept jobs outside the private law firm industry. Further, both groups reduce their hours even though their wage rates appear to be constant or increasing. Among those who make partner, we do not see significant changes in hours worked.

We can easily pick parameters for our model that produce both of these patterns. In our model, associates take on heavy workloads as part of an audition process, and once their auditions are over, they no longer earn the same information returns from their work. This is true for those who make partner as well, but those who earn partner positions have additional incentives to keep working long hours because they begin taking on valuable leadership and business development roles that increase their productivity.

Before describing our results, we must discuss several issues that arise concerning how to map our results into data. In our model, there are only two positions, associate and partner. For much of the 20th century, most private law firms created only these two positions, but in recent decades, many firms have created additional positions. Many firms now have non-equity partner positions, and a smaller number have "Of Counsel" or counsel positions.

We focus on lawyers who are in or just beyond the first decade of their careers, and based on our reading of the literature and our work with the panel data set described below, we proceed under the assumption that persons who are promoted from associate to non-equity partner early in their careers are likely persons who are still trying to earn promotions to full partner. We assume that persons who transition from associate to counsel positions early in their careers are no longer being considered for promotion to partner.[27]

Appendix D demonstrates that non-equity partner can be a stepping stone between associate and partner. However, as a rule, new associates who move into counsel positions do not become partners. Given these patterns, we assume that

[27]Among more experienced lawyers, some counsels and non-equity partners are former partners in their firms or other firms who voluntarily or involuntarily left their partnership positions because they were unwilling or unable to meet the expectations of other partners. See Richmond [2010].

lawyers who begin their careers as associates but transition into counsel positions are persons who learned that they were not well suited to the partner role.[28] In addition, both they and the market learned over time that they are particularly well-suited to a senior, non-partner role that likely requires special expertise.

Given this scenario, consider Table 2. This table describes data from the Survey of Law Firm Economics (SLFE), which is conducted annually by ALM Legal Intelligence. The data come from eight annual surveys taken during the period 2007-2014. Some firms appear in more than one annual survey, but these are not observations from a panel data set. Rather, the data come from eight repeated cross-sectional surveys.[29]

Table 2 describes outcomes for lawyers who are between eight and twelve years into their careers. We chose this experience interval because law firms now make many crucial retention and promotion decisions in this interval. The table presents results for associates, equity partners, non-equity partners, and counsel attorneys. Because these counsel attorneys are less than 12 years into their careers, it seems reasonable to assume that most of them are attorneys who recently left an associate position in their current or previous firms and now occupy a senior role off the partnership track. Thus, it is interesting to note that, compared to the other three groups, counsel attorneys bill fewer hours but charge clients higher rates for their time. It is particularly noteworthy that counsels bill their time at rates 17 percent greater than associates yet bill 24 percent fewer hours.[30]

These data contain information about hours billed but do not measure total hours worked. Therefore, we cannot use these data to recover actual wage rates for workers in different positions. Still, given the large gaps in hours billed and billing rates, it seems highly unlikely that the associates in Table 2 are earning higher hourly wages than counsel attorneys who have comparable total experience. So, why would these associates bill many more hours than their peers who work as counsels? Our model suggests that this hours gap reflects the fact that counsel attorneys are no longer auditioning for partnerships. Their past work experience has revealed their type, and they are no longer producing signals about their suitability for the partner position.

Some readers may conjecture that these results for counsel attorneys are a "mommy track" outcome, but these results are not driven by women taking counsel positions in order to spend more time with young children. We have created a similar table that contains only male lawyers, and the results are quite similar.[31]

[28]Also, see Scharf and Flom [2011] page 9. Using different data, they also conclude that counsel positions are "off-track."

[29]Some lawyers may appear in two different cross-sections, but we cannot link these records.

[30]Given the experience restrictions we impose on our sample, the associates in Table 2 are persons who are reaching the end of their tenure as associates, and in some cases, may already know that they are not going to be promoted and are therefore engaged in searches for new positions. Nonetheless, these associates still bill more hours than counsel attorneys who have similar levels of experience and bill their time at higher rates.

[31]See Harrington and Hsi [2007] for a discussion of links between gender differences in promotion rates and gender differences in time spent caring for young children. Gender is missing for many records in the SLFE data. However, in the sample of respondents who report being male, the same patterns exit, and the hours billed gap between Associates and Counsels is even slightly larger than the one reported in Table 2 for the full sample.

Table 2
Hours Billing, Billing Rates, &
Total Compensation By Position

	Partner	NE Partner	Associate	Counsel
Avg Hours Billed	1,646	1,612	1,493	1,128
(Std Deviation)	(416)	(538)	(592)	(633)
N	2,982	3,135	7,144	558
Avg Hourly Rate	290	299	259	304
(Std Deviation)	(77)	(97)	(85)	(105)
N	2,990	3,131	6,931	543
Avg Comp	233,970	197,471	143,409	144,305
(Std Deviation)	(110,968)	(78,307)	(53,157)	(67,458)
N	3,053	3,188	7,283	574

Notes: This table reports job characteristics for lawyers included in the Survey of Law Firm Economics taken by ALM Legal Intelligence between the years 2007 and 2014. Within each panel of this table, the first number in each cell is the sample mean, the numbers in parentheses are standard deviations, and N is the sample size. This table considers only attorneys with between 8 and 12 years of experience, defined as years since passing the bar. Compensation is defined as take home salary and retirement contributions plus year end bonus plus benefits. Hours Billed is the annual number of billable hours for each attorney. Hourly rate is the typical hourly rate charged by each attorney. Sample sizes differ among cells due to different frequencies of item non-response.

Partners and non-equity partners also work more than counsel attorneys, but this pattern is easy to understand. The implied hourly wage rate for new partners may be much greater than the implied wage rate for counsels, and many non-equity partners are still auditioning for full partnerships, receiving performance-based profit sharing, or both.[32]

The results in Table 2 are only suggestive because they do not contain panel data on particular lawyers and therefore do not allow us to know with certainty that these counsel attorneys recently left associate positions. Tables 3 and 4 present results from a second data set that provides much smaller samples but does allow us to track individual lawyers over time.

[32]We have created a version of Table 2 that restricts the sample to lawyers who work in firms with at least 150 attorneys. Among this group of firms, partner positions are more homogeneous in terms of status and earnings potential. Within this sample, the patterns we highlight in Table 2 remain, but hours-billed, billing-rates, and total compensation are higher within all job categories.

Table 3
AJD Full Sample
Hours Worked and Compensation By W3 Position

	Partner	NE Partner	Associate	Counsel	Other
Avg Hours W1	51.7	51.8	49.5	49.7	49.9
(Std Deviation)	(12.7)	(9.4)	(12.4)	(11.2)	(13.1)
N	161	103	75	31	343
Avg Salary W1	116,243	122,125	92,157	141,578	131,753
(Std Deviation)	(48,985)	(48,303)	(39,828)	(47,452)	(70,311)
N	165	107	74	32	360
Avg Hours W3	51.2	54.3	51.9	46.2	47.0
(Std Deviation)	(12.8)	(12.2)	(10.7)	(12.2)	(12.8)
N	171	109	77	32	365
Avg Comp W3	240,154	225,653	134,187	215,975	190,566
(Std Deviation)	(220,199)	(110,588)	(75,675)	(136,151)	(197,826)
N	121	90	72	26	316

Notes: This table reports job characteristics for lawyers included in the After the JD Survey (AJD), which was sponsored by the American Bar Association. This panel survey involved three rounds of data collection that began in 2002, 2007, and 2012. The first number in each cell is the sample mean, the numbers in parentheses are standard deviations, and N is the sample size. The sample for this table includes only attorneys who passed the bar in 1998 or later and were associates at private law firms in the first wave of the survey. The columns group respondents by their Wave 3 employment status. The first four columns contain persons who are employed in private law firms in Wave 3. We eliminate five respondents who report working in staff or contract positions. The Other category includes solo practitioners and persons employed by an organization that is not a private law firm. Average Hours Wave 1 and Average Hours Wave 3 are average hours worked by attorneys in the weeks before completing the first and third surveys, respectively. Average Salary Wave 1 is defined as the salary including bonus. Average Compensation Wave 3 is the average sum of salary, bonus, profit sharing, and other income received by respondents in the third wave. Sample sizes differ by cell due to different frequencies of item non-response.

The *After the JD* (AJD) study conducted three rounds of interviews with a cohort of lawyers who passed the bar around 2000. Wave one interviews took place between May, 2002 - March, 2003. Wave two interviews began in May, 2007 and ran through early 2008. Wave 3 interviews took place between May, 2012 and December, 2012. We chose a subsample of these lawyers who participated in both the wave 1 and wave 3 interviews and who reported in wave 1 that they worked as an associate in a private law firm. By wave 3, most of these lawyers should have had eleven or twelve years of experience as lawyers. Thus, they are slightly more experienced than the average lawyer in our Table 2 sample.

Each column in Table 3 describes outcomes for lawyers who were associates in wave 1 and occupy a specific position in wave 3. As in Table 2, we present results for equity partners, non-equity partners, associates, and counsels, but we also present an "other" category. This category contains almost half of our sample and includes lawyers who have left private law practice or are self-employed.

The first row of Table 3 reports hours worked in wave 1. All of the lawyers in this sample were working as associates in private law firms in wave 1. Young associates who are going to make partner or non-equity partner appear to work slightly longer hours in wave 1. Gicheva [2013] notes that, if young workers differ in their costs of effort and also learn by doing, those with low costs of effort work more and are more likely to acquire skills that earn promotions. Further, since the wave 1 interviews took place in year two or three of these young lawyers' careers, the small differences in wave 1 hours worked recorded in the first row of Table 3 may indicate that partners give more work to second and third year associates who, based on the quality of their early work, appear more likely to make partner.[33]

In wave 3, we do see more noteworthy differences in hours worked among these lawyers. The small number who took counsel positions and the much larger sample who no longer work in private law firms work about three hours less per week than they worked as new associates. In contrast, those in associate and non-equity partner positions who are still trying to become equity partners work more than they worked in wave 1, while those who are partners in wave 3 work thirty minutes less per week than they worked in wave 1.

The wave 3 contrast between associates and lawyers in our other category is striking. In wave 1, when both groups were working as associates, those in the other category enjoyed higher annual earnings and worked a few minutes more per week. Between waves 1 and 3, annual earnings grew for both groups by about 45 percent, but in wave 3, those who remain in associate positions work roughly five hours per week more than those who have left private law. The results for the small sample of attorneys who are counsels in wave 3 parallel those for attorneys in the other category, except they enjoy slightly higher earnings and earnings growth while reporting slightly larger declines in hours between wave 1 and wave 3.

Taken as a whole, the wave 3 comparisons between those who are partners or still trying to become partners and those who have left the partnership track provide additional support for our claim that new associates work long hours as part of a screening process. When new associates learn they are not going to make partner, the information value produced by their effort is diminished, and they reduce their hours even though their wage rates appear to be constant or rising. Those who make partner have also completed the screening process, but they now take on new leadership and business development roles that raise the value of their work effort.

Table 4 summarizes the association between leaving the partnership track and changes in hours worked between waves 1 and 3 for several different samples. Here, we define terms as follows: within our sample of wave 1 associates, those who are partners, non-equity partners, or associates in wave 3 are still on the partnership track, but self-employed lawyers, counsels, and lawyers who have left private practice are off the partnership track. Table 4 reports the results from a series of bivariate regressions of hours worked in wave 3 minus hours worked in wave 1 on a

[33]Wilkins and Gulati [1998] discuss differences in work assignments among associates.

dummy variable indicating whether or not a lawyer is off the partnership track in wave 3.

Lawyers who leave the partnership track reduce their hours by roughly 3 to 7 hours per week relative to lawyers who remain on the partnership track. The differences among the estimated effects in the five columns are not statistically significant, but there is some indication that the effect is stronger among lawyers who began their careers in large private law firms. There is no evidence that this result is driven purely by women seeking to reduce their hours in order to spend more time at home. The estimated effect among women is slightly larger in the full sample but smaller in the sample of lawyers who began their careers at law firms with more than 150 attorneys.

Table 4
Regressions of Changes in Hours (W3 - W1)
on an Indicator for Leaving Partnership Track

$Y = \Delta$Hours (W3 - W1) $X = 1$ if Off-Track in W3 (Counsel, Other)

	Full	M	F	M Firm Size = 150+	F Firm Size = 150+
Off-Track W3	-3.70	-3.25	-4.26	-7.43	-5.01
(Std Error)	(1.15)	(1.53)	(1.78)	(2.76)	(2.92)
N	708	425	279	178	114

Notes: See notes to Table 3 for description of the AJD sample. Once again, we restrict our samples to persons who report in Wave 1 that they are associates in private law firms, and we eliminate five respondents who report working in staff or contract positions in Wave3. Those who are Equity Partners, Non-equity Partners, or Associates in a private law firm are On-Track. Those who are solo practitioners, counsels in a private law firm or employees of an organization that is not a private law firm are Off-Track. The entries here are regression coefficients on a dummy variable indicating that, in Wave 3, a lawyer is Off-Track. All regressions are bivariate regressions of changes in hours worked per week, i.e. wave 3 hours minus wave 1 hours, on the Off-Track indicator. Standard errors are in parentheses. N denotes sample size.

The empirical results in Tables 2, 3, and 4 are only suggestive. The AJD surveys have response rates far below one, and the SLFE documentation does not report response rates. Further, neither data source allows us to measure changes in wage rates and hours worked that occur exactly when lawyers make specific job or career changes.

Nonetheless, the data do suggest that many lawyers who abandon the partnership track in private law firms reduce their work hours even though their wage rates are constant or rising. Because these changes in work hours accompany changes in job assignments, it is reasonable to suspect that these workers are adjusting their hours in response to changes in their information sets. Our model suggests that they adjust their hours because they have recently learned that they are not well-suited to the partner role and that the long hours they worked as new associates were part of the audition process for partnerships.

These patterns and other features of the data are not easy to generate in a model where associates work more than the efficient number of hours simply to signal their private information about their effort costs. Promotion rates are low in these markets[34] and lawyers in the AJD do not cut back significantly on their hours after they become partners, as we would expect if the long hours that associates work are pure signals of their effort costs.

8. Conclusion

We argue that new associates in professional service firms take on heavy workloads while simultaneously competing in up-or-out promotion contests because both practices facilitate the identification of the new partners that these firms need to survive and grow. Partners in professional service firms possess a rare combination of skills. They possess the analytical skills needed to perform and direct complex work, and they possess the communication and people skills required to earn and maintain the trust of valuable clients. Further, because the trust relationship between a partner and her clients hinges on the partner's promise to reliably provide expert services, each partner can only manage a limited number of clients.[35] Given this constraint, professional service firms grow horizontally, by identifying new partners who can build and maintain relationships with new clients.

Young professionals in elite professional service firms take on heavy workloads so that both they and their employers better learn whether or not they should be new partners. Those who discover that they are not going to become partners are no longer willing to bear these workloads, and it is efficient for firms to replace them with new associates who are eager to discover whether or not they can become partners.

In recent decades, mergers have created some large law and public accounting firms that no longer adhere strictly to up-or-out rules. These firms have created a limited number of productive positions for specialists who may not be well-suited to the partner role but are able to provide expert services that multiple partners employ. Future research should more closely examine how recent developments in information technology and the growth of large firms have shaped personnel policies within professional service industries. Nonetheless, because professional service firms remain dependent on partners to develop and maintain business, most new entrants in these firms still begin their careers believing that they are engaged in auditions for partnerships.

[34]The vast majority of Wave 1 associates in the AJD leave their Wave 1 firm before Wave 3, and less than one in ten make partner in their Wave 1 firm by Wave 3.

[35]See Levin and Tadelis [2005] for a model that explores why corporate clients are not willing to have professionals who are not partners manage their cases. In contrast, many personal injury law firms do not generate revenue by building and maintaining relationships with corporate clients. They generate business through advertising and mass marketing. Galanter and Palay [1998] argue that this may explain why personal injury firms are often not organized around the traditional associate to partner career path.

References

George Akerlof. The economics of caste and of the rat race and other woeful tales. *The Quarterly Journal of Economics*, 90(4):599–617, 1976.

Elizabeth Dreike Almer. *AICPA Work/Life and Women's Initiatives 2004 Research: Executive Summary*, 2004.

Charles Batchelor. Up or out is part of industry culture. *Financial Times*, April 20, 2011.

Marianne Bertrand, Claudia Goldin, and Lawrence Katz. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economics Journal: Applied Economics*, 2:228–255, July 2010.

Alessandro Bonatti and Johaness Horner. Career concerns with exponential learning. *Theoretical Economics*, 12(1):425–475, 2017.

Gautam Bose and Kevin Lang. Monitoring for worker quality. *Journal of Labor Economics*, 35(3), 2017.

Dominique Demougin and Aloysius Siow. Careers in ongoing hierarchies. *The American Economic Review*, 84(5):1261–1277, 1994.

Marc Galanter and Thomas Palay. *Tournament of Lawyers: The Transformation of the Big Law Firm*. The University of Chicago Press, Chicago, 1991.

Marc Galanter and Thomas Palay. A little jousting about the big law firm tournament. *Virginia Law Review*, 84(8):1683–1693, 1998.

Marc S Galanter and William D. Henderson. The elastic tournament: A second transformation of the big law firm. *Stanford Law Review*, 60(1867):1867–1930, 2008.

Suman Ghosh and Michael Waldman. Standard promotion practices versus up-or-out contracts. *The RAND Journal of Economics*, 41(2):301–325, 2010.

Robert Gibbons and Michael Waldman. A theory of wage and promotion dynamics. *Quarterly Journal of Economics*, 114(4):1321–1358, 1999.

Robert Gibbons and Michael Waldman. Enriching a theory of wage and promotion dynamics inside firms. *Journal of Labor Economics*, 24(1):59–107, 2006.

Dora Gicheva. Working long hours and early career outcomes in the high-end labor market. *Journal of Labor Economics*, 31(4):785–824, 2013.

Ronald J. Gilson and Robert H. Mnookin. Sharing among the human capitalists: An economic inquiry into the corporate law firm and how partners split profits. *Stanford Law Review*, 37(2):313–392, 1985.

Ronald J. Gilson and Robert H. Mnookin. Coming of age in a corporate law firm: The economics of associate career patterns. *Stanford Law Review*, 41(3):567–595, 1989.

Claudia Goldin. A grand gender convergence: Its last chapter. *American Economic Review*, 104(4):1091–1119, 2014.

Elizabeth H. Gorman. Moving away from up-or-out: Determinants of permanent employment in law firms. *Law and Society Review*, 33(3):637–666, 1999.

Steven Greenhouse. Flex time flourishes in accounting industry. *The New York Times*, jan 2011.

Mona Harrington and Helen Hsi. Women lawyers and obstacles to leadership. MIT Workplace Center, 2007.

Bengt Holmstrom. Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, 66(1):169 – 182, 1999.

Charles Kahn and Gur Huberman. Two-sided uncertainty and up-or-out contracts. *Journal of Labor Economics*, 6(4):423–444, 1988.

Renee M. Landers, James B. Rebitzer, and Lowell J. Taylor. Rat race redux: Adverse selection in the determination of work hours in law firms. *The American Economic Review*, 86(3):329–348, 1996.

Jonathan Levin and Steven Tadelis. Profit sharing and the role of professional partnerships. *Quarterly Journal of Economics*, pages 131–171, 2005.

John Lewison. The work/life balance sheet so far. *Journal of Accountancy*, aug 2006.

Brendan O'Flaherty and Aloysius Siow. On the job screening, up or out rules, and firm growth. *The Canadian Journal of Economics*, 25(2):346–368, 1992.

Canice Prendergast. The role of promotion in inducing specific human capital acquisition. *The Quarterly Journal of Economics*, 108(2):523–534, 1993.

Michael J. Pries. Persistence of employment fluctuations: A model of recurring job loss. *The Review of Economics Studies*, 71(1):193–215, 2004.

James Rebitzer and Lowell Taylor. When knowledge is an asset: Explaining the organizational structure of large law firms. *Journal of Labor Economics*, 25(2): 201–229, 2007.

Douglas Richmond. The partnership paradigm and law firm non-equity partners. *Kansas Law Review*, 58:507 – 551, 2010.

Sherwin Rosen. Learning and experience in the labor market. *Journal of Human Resources*, 7(3):326 – 342, 1972.

Stephanie A. Scharf and Barbara M. Flom. Report of the fifth annual national survey on retention and promotion of women in law firms. National Association of Women Lawyers and the NAWL Foundation, oct 2011.

Michael Waldman. Up-or-out contracts: A signaling perspective. *Journal of Labor Economics*, 8(2):230–250, 1990.

David B. Wilkins and G. Mitu Gulati. Reconceiving the tournament of lawyers: Tracking, seeding, and information control in the internal labor markets of elite law firms. *Virginia Law Review*, 84:1581–1681, 1998.

Appendix A: Planner's Problem and Proofs

In this Appendix, we provide the proofs to claims made in the paper. Recall that the planner's problem is given by the Bellman equation

$$V(\rho^u,\rho^x) = \max_{\alpha^x,\alpha^u,n} s(\alpha^u,\alpha^x,n) + \beta V\left(\rho^u_{+1},\rho^x_{+1}\right) \tag{8.1}$$

where

$$s(\alpha^u,\alpha^x,n) \equiv (2-q)\,v^o + q v^a_0(n) + \alpha^x\rho^x\left(v^p_1 - v^a_0(n)\right) + \alpha^u\rho^u\left(v^a_1 - v^a_0(n)\right)$$

The laws of motion for the state variables ρ^u and ρ^x are given by

$$\begin{aligned} \rho^u_{+1} &= (q-\rho^u\alpha^u-\rho^x\alpha^x)(1-\phi(n)) \\ \rho^x_{+1} &= (q-\rho^u\alpha^u-\rho^x\alpha^x)\,\pi\phi(n) \end{aligned}$$

Finally, the planner faces the constraint

$$q-\rho^u\alpha^u-\rho^x\alpha^x \geq 0 \tag{8.2}$$

We begin with some preliminary results presented in the text.

Claim 1: $V(\rho^u,\rho^x) = K_1 + K_2\rho^u + K_3\rho^x$ for all (ρ^u,ρ^x) s.t. $0 \leq \rho^u+\rho^x < q$. Moreover, the optimal plan ensures $0 \leq \rho^u_t+\rho^x_t < q$ for all $t \geq 1$ starting from any initial condition (ρ^u_0,ρ^x_0).

Proof of Claim 1: Since α^u and α^x are at most 1, constraint (8.2) will be satisfied whenever $\rho^u+\rho^x < q$, i.e. whenever the total number of experienced workers is less than the number of jobs in the professional sector. We first restrict attention to this case, and then argue that the optimal plan ensures $0 \leq \rho^u_t+\rho^x_t < q$ for all $t \geq 1$ starting at any initial condition (ρ^u_0,ρ^x_0).

The Bellman equation (8.1) can be written as a functional equation

$$V = T(V)$$

where T is an operator defined over the space of bounded functions V that map the domain $\{(\rho^u,\rho^x) : 0 \leq \rho^u+\rho^x \leq 1\}$ into $\mathbb{R}$.

First, we argue that T is a contraction, i.e. for any two functions V_1 and V_2 that map $\{(\rho^u,\rho^x) : 0 \leq \rho^u+\rho^x \leq 1\}$ into $\mathbb{R}$, $||T(V_1)-T(V_2)|| < ||V_1-V_2||$ where the distance between functions is defined as

$$||V_2-V_1|| = \sup_{\{(\rho^u,\rho^x):0\leq\rho^u+\rho^x\leq 1\}} |V_2(\rho^u,\rho^x) - V_1(\rho^u,\rho^x)|$$

To verify T is a contraction, it suffices to verify Blackwell's sufficient conditions. In stating these, we adopt the convention that a function $V_1 \leq V_2$ if $V_1(\rho^u,\rho^x) \leq V_2(\rho^u,\rho^x)$ for all $\{(\rho^u,\rho^x) : 0 \leq \rho^u+\rho^x \leq 1\}$.

(1) Monotonicity: If $V_1 \leq V_2$, then $T(V_1) \leq T(V_2)$
(2) Discounting: there exists some $\beta \in (0,1)$ such that $T(V(\rho^u,\rho^x)+a) \leq T(V(\rho^u,\rho^x)) + \beta a$ for all $a \geq 0$ for all (ρ^u,ρ^x)

Both of these are straightfoward to verify. Define

$$\begin{aligned} \overline{T}(V;\alpha^u,\alpha^x,n) &= v^o + v^a_0(n) + \alpha^u\rho^u\left(v^a_1 - v^a_0(n)\right) + \alpha^x\rho^x\left(v^p_1 - v^a_0(n)\right) + \\ &\quad \beta V\left((q-\rho^u\alpha^u-\rho^x\alpha^x)(1-\phi(n)),\,(q-\rho^u\alpha^u-\rho^x\alpha^x)\,\pi\phi(n)\right) \end{aligned}$$

i.e. $\overline{T}(V)$ evaluates V for a fixed vector (α^u,α^x,n) rather than for the vector that maximizes the RHS of (8.1) subject to constraints. Let $(\hat{\alpha}^u,\hat{\alpha}^x,\hat{n})$ be the values that solves the RHS of (8.1) when $V = V_1$, and which are functions of (ρ^u, ρ^x). If $V_2 \geq V_1$, then we then have

$$\begin{aligned} T(V_2) &\geq \overline{T}(V_2; \hat{\alpha}^u, \hat{\alpha}^x, \hat{n}) \\ &\geq \overline{T}(V_1; \hat{\alpha}^u, \hat{\alpha}^x, \hat{n}) \\ &= T(V_1) \end{aligned}$$

This establishes monotonicity. For discounting, observe that replacing V with $V + a$ will leave the arg max on the RHS of (8.1) unchanged. Hence,

$$T(V_1 + a) = T(V_1) + \beta a$$

where β is the discount rate and thus less than 1. It follows that T is a contraction. Hence, there exists a unique fixed point V in the set of bounded functions such that $V = T(V)$.

Next, we argue that V is linear in ρ^u and ρ^x over the set of all (ρ^u, ρ^x) for which $0 \leq \rho^u + \rho^x < q$. To prove this, it will be enough to show that if V is linear over this region, then $T(V)$ must be linear over this region as well. In that case, the contraction mapping theorem tells us there exists a fixed point within the set of functions V that are linear over this region, ensuring V is linear in this region.

Consider a function $V(\rho^u, \rho^x)$ that is linear over the set of all (ρ^u, ρ^x) for which $0 \leq \rho^u + \rho^x < q$, i.e. for functions that have the form

$$V(\rho^u, \rho^x) = K_1 + K_2 \rho^u + K_3 \rho^x \tag{8.3}$$

when $0 \leq \rho^u + \rho^x < q$. Since $v_0^a(n)$ and $\phi(n)$ are both concave in n, the RHS of (8.1) is concave in n when $V(\rho^u, \rho^x)$ is linear. Hence, for any value of α^x and α^u, the necessary and sufficient condition for $\hat{n}$ to be optimal is

$$(q - \rho^u \alpha^u - \rho^x \alpha^x) \left. \frac{dv_0^a}{dn} \right|_{n=\hat{n}} + (q - \rho^u \alpha^u - \rho^x \alpha^x) \beta \phi'(n) (\pi K_3 - K_2) = 0$$

Since $\rho^u + \rho^x < q$, we have $q - \rho^u \alpha^u - \rho^x \alpha^x > 0$, so we can divide the above equation by $(q - \rho^u \alpha^u - \rho^x \alpha^x)$. This leaves us with the first order condition

$$(1 + \pi x) - c'(\hat{n}) + \beta \phi'(n) (\pi K_3 - K_2) = 0$$

It follows that the optimal $\hat{n}$ which maximizes (8.1) is independent of ρ^u and ρ^x whenever $0 \leq \rho^u + \rho^x < q$, although it will depend on the coefficients K_2 and K_3. Next, since the objective function above is linear in α^u and α^x, we can deduce that the following scheme is optimal:

$$\hat{\alpha}^u = \begin{cases} 1 & \text{if } v_1^a - v_0^a - \beta [\pi \phi(n) K_3 + (1 - \phi(n)) K_2] \geq 0 \\ 0 & \text{if } v_1^a - v_0^a - \beta [\pi \phi(n) K_3 + (1 - \phi(n)) K_2] < 0 \end{cases} \tag{8.4}$$

$$\hat{\alpha}^x = \begin{cases} 1 & \text{if } v_1^p - v_0^a - \beta [\pi \phi(n) K_3 + (1 - \phi(n)) K_2] \geq 0 \\ 0 & \text{if } v_1^p - v_0^a - \beta [\pi \phi(n) K_3 + (1 - \phi(n)) K_2] < 0 \end{cases} \tag{8.5}$$

Note that when the RHS of the above equations is exactly equal to 0, any value of α^u or α^x yields the same value for the objective function. We adopt the convention of setting $\hat{\alpha}^u$ and $\hat{\alpha}^x$ equal to 1 in these cases only for notational convenience. Since $\hat{n}$ and v_0^a are independent of ρ^u and ρ^x when $0 \leq \rho^u + \rho^x < q$, it follows that the optimal $\hat{\alpha}^u$ and $\hat{\alpha}^x$ are independent of (ρ^u, ρ^x) in this region as well. As a result, the fact that V is linear over a given region implies that $T(V)$ is linear over the same region as well, i.e.

$$T(V) = \overline{K}_1 + \overline{K}_2\rho^u + \overline{K}_3\rho^x$$

for $0 \leq \rho^u + \rho^x < q$, where

$$\begin{aligned}
\overline{K}_1 &= v^o + v_0^a + \beta q\left[\pi\phi(\hat{n}) K_3 + (1-\phi(\hat{n})) K_2\right] + \beta K_1 \\
\overline{K}_2 &= \hat{\alpha}^u\left(v_1^a - v_0^a\right) - \beta\hat{\alpha}^u\left[\pi\phi(\hat{n}) K_3 + (1-\phi(\hat{n})) K_2\right] \\
\overline{K}_3 &= \hat{\alpha}^x\left(v_1^p - v_0^a\right) - \beta\hat{\alpha}^x\left[\pi\phi(\hat{n}) K_3 + (1-\phi(\hat{n})) K_2\right]
\end{aligned}$$

Note that the coefficients $\overline{K}_1$, $\overline{K}_2$, and $\overline{K}_3$ that define $T(V)$ are not simple linear expressions of K_1, K_2, and K_3, since $\overline{K}_1$, $\overline{K}_2$, and $\overline{K}_3$ depend on $\hat{n}$, $\hat{\alpha}^u$, and $\hat{\alpha}^x$, and these are all non-linear functions of K_1 and K_2.

Finally, observe that if we add up the laws of motion for ρ_t^u and ρ_t^x, we obtain

$$\rho_{t+1}^u + \rho_{t+1}^x = \left(q - \rho_t^u\hat{\alpha}^u - \rho_t^x\hat{\alpha}^x\right)\left(1 - (1-\pi)\phi(\hat{n})\right) \tag{8.6}$$

Note that both expressions on the RHS of (8.6) are nonnegative. The first expression is nonnegative from (8.2), while the second is positive given our assumptions that the maximal effort level $\overline{n}$ is such that $\phi(\overline{n}) < 1$. It follows that $\rho_{t+1}^u + \rho_{t+1}^x \geq 0$. For a given (ρ_t^u, ρ_t^x), we consider two possible cases. First, suppose $\rho_t^u\hat{\alpha}^u + \rho_t^x\hat{\alpha}^x > 0$. In this case, the first term in (8.6) will be less than q, while the second term in (8.6) will at most 1. Hence, $\rho_{t+1}^u + \rho_{t+1}^x < q$. Next, suppose $\rho_t^u\hat{\alpha}^u + \rho_t^x\hat{\alpha}^x = 0$. In this case, young associates will be employed at date t and asked to put in a positive amount of effort, in which case the optimal effort of young workers will satisfy $\hat{n}_t > 0$. In this case, the first term in (8.6) will be at most q while the second term in (8.6) will be strictly less than 1. In that case, we would once again have $\rho_{t+1}^u + \rho_{t+1}^x < q$. By induction, we can conclude that regardless of the initial condition (ρ_0^u, ρ_0^x), the optimal path will imply $\rho_t^u + \rho_t^x < q$ for all $t \geq 1$. ■

Claim 2: The optimal value $\hat{\alpha}^x = 1$ for any pair (ρ^u, ρ^x) s.t. $0 \leq \rho^u + \rho^x < q$.

Proof of Claim 2: Using Claim 1, we can write $V(\rho^u, \rho^x)$ whenever $0 \leq \rho^u + \rho^x < q$ as

$$V(\rho^u, \rho^x) = K_1 + K_2\rho^u + K_3\rho^x$$

Matching coefficients, K_1, K_2, and K_3 must satisfy the following for the Bellman equation to be satisfied for values of (ρ^u, ρ^x) such that $0 \leq \rho^u + \rho^x < q$:

$$K_1 = (2-q)v^o + qv_0^a + \beta q\left[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2\right] + \beta K_1 \tag{8.7}$$

$$K_2 = \hat{\alpha}^u\left(v_1^a - v_0^a\right) - \beta\hat{\alpha}^u\left(\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2\right) \tag{8.8}$$

$$K_3 = \hat{\alpha}^x\left(v_1^p - v_0^a\right) - \beta\hat{\alpha}^x\left(\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2\right) \tag{8.9}$$

Using equations (8.8) and (8.9) to solve for K_2 and K_3 yields

$$K_2 = \frac{\hat{\alpha}^u\left(v_1^a - v_0^a\right) - \beta\hat{\alpha}^u\hat{\alpha}^x\pi\phi(\hat{n})\left(v_1^p - v_1^a\right)}{1 + \beta\hat{\alpha}^u(1-\phi(\hat{n})) + \beta\hat{\alpha}^x\pi\phi(\hat{n})} \tag{8.10}$$

$$K_3 = \frac{\hat{\alpha}^x\left(v_1^p - v_0^a\right) + \beta\hat{\alpha}^u\hat{\alpha}^x(1-\phi(\hat{n}))\left(v_1^p - v_1^a\right)}{1 + \beta\hat{\alpha}^u(1-\phi(\hat{n})) + \beta\hat{\alpha}^x\pi\phi(\hat{n})} \tag{8.11}$$

Note that K_2 and K_3 are both non-negative. The first order condition for $\hat{\alpha}^x$, in line with (8.5), implies $\hat{\alpha}^x = 1$ whenever

$$v_1^p - v_0^a \geq \beta\left(\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2\right)$$

Since $\beta < 1$, it will suffice to show that K_2 and K_3 are bounded above by $v_1^p - v_0^a$, since this would imply

$$\beta \left(\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))\, K_2\right) \leq \beta \left(v_1^p - v_0^a\right) < v_1^p - v_0^a$$

Begin with K_2. Observe that

$$\begin{aligned} v_1^p &\equiv \max_n\,(1+x)\, z^p n - c\,(n) \\ &\geq \max_n\,(1+\pi x)\, z^a n - c\,(n) \equiv v_1^a \end{aligned}$$

Since $v_1^p - v_1^a \geq 0$, we have

$$\begin{aligned} K_2 &= \frac{\hat{\alpha}^u \left(v_1^a - v_0^a\right) - \beta\hat{\alpha}^u\hat{\alpha}^x\pi\phi(\hat{n})\left(v_1^p - v_1^a\right)}{1 + \beta\hat{\alpha}^u\left(1-\phi(\hat{n})\right) + \beta\hat{\alpha}^x\pi\phi(\hat{n})} \\ &\leq \frac{\hat{\alpha}^u \left(v_1^a - v_0^a\right)}{1 + \beta\hat{\alpha}^u\left(1-\phi(\hat{n})\right) + \beta\hat{\alpha}^x\pi\phi(\hat{n})} \\ &\leq v_1^a - v_0^a \\ &\leq v_1^p - v_0^a \end{aligned}$$

Next, consider K_3. Observe that

$$\begin{aligned} v_1^a &\equiv \max_n\,(1+\pi x)\, z^a n - c\,(n) \\ &\geq \max_n\,(1+\pi x)\, n - c\,(n) \geq v_0^a \end{aligned}$$

This implies $v_1^p - v_0^a \geq v_1^p - v_1^a$, and so

$$\begin{aligned} K_3 &= \frac{\hat{\alpha}^x \left(v_1^p - v_0^a\right) + \beta\hat{\alpha}^u\hat{\alpha}^x\left(1-\phi(\hat{n})\right)\left(v_1^p - v_1^a\right)}{1 + \beta\hat{\alpha}^u\left(1-\phi(\hat{n})\right) + \beta\alpha^x\pi\phi(\hat{n})} \\ &\leq \frac{1 + \beta\hat{\alpha}^u\left(1-\phi(\hat{n})\right)}{1 + \beta\hat{\alpha}^u\left(1-\phi(\hat{n})\right) + \beta\hat{\alpha}^x\pi\phi(\hat{n})}\hat{\alpha}^x\left(v_1^p - v_0^a\right) \\ &\leq v_1^p - v_0^a \end{aligned}$$

It follows that $\hat{\alpha}^x = 1$ for these values of (ρ^u, ρ^x) ■

Lemma 1: $\pi K_3 > K_2$, where K_2 and K_3 are the coefficients of the value function for (ρ^u, ρ^x) s.t. $0 \leq \rho^u + \rho^x < q$.

Proof of Lemma 1: Consider the expression for K_1 as defined in equation (8.7). This corresponds to the present discounted surplus the planner can expect to generate starting with no experienced workers, i.e. starting with $\rho^u = \rho^x = 0$. Since the planner always assigns $2-q$ workers to the outside sector, and they each produce a surplus of v^o. The present discounted value of this surplus is just $(1-\beta)^{-1}(2-q)\, v^o$. Hence, the difference $K_1 - (1-\beta)^{-1}(2-q)\, v^o$ is equal to the net surplus the planner expects to generate in the professional sector when $\rho^u = \rho^x = 0$, i.e. it is equal to the value of staffing all q positions in the professional sector with young workers, observing their output, and then staffing these positions optimally thereafter. Since information on worker types can be used to set their hours optimally, it follows that this value must be strictly larger than the surplus generated in the professional sector from staffing all q positions in the professional sector with young workers, ignoring any information that may be revealed about their quality and treating all such workers as if their ability was uncertain, but acting optimally in using information thereafter. The latter yields a value of

$$qv_0^a + \beta\left(K_1 - (1-\beta)^{-1}(2-q)\, v^o + qK_2\right)$$

In particular, $K_1 - (1-\beta)^{-1}(2-q)v^o$ represents the value of staffing all professional jobs with young workers, and qK_2 represents the incremental value of having a mass q of experienced workers of uncertain ability that can be employed in the professional sector. Since using the information is more valuable, we have

$$K_1 - (1-\beta)^{-1}(2-q)v^o > qv_0^a + \beta\left(K_1 - (1-\beta)^{-1}(2-q)v^o + qK_2\right)$$

which after rearranging yields

$$(1-\beta)K_1 > (2-q)v^o + qv_0^a + \beta qK_2 \tag{8.12}$$

From equation (8.7) above we know that K_1 satisfies

$$(1-\beta)K_1 = (2-q)v^o + qv_0^a + \beta q\pi\phi(\hat{n})K_3 + \beta q(1-\phi(\hat{n}))K_2$$

Rearranging this equation implies

$$\beta\phi(\hat{n})(\pi K_3 - K_2) = (1-\beta)K_1 - (2-q)v^o - qv_0^a - \beta qK_2 \tag{8.13}$$

The RHS of (8.13) is strictly positive given (8.12). Since it will always be optimal to have inexperienced workers put in some effort, $\phi(\hat{n}) > 0$. Hence, $\beta\phi(\hat{n})(\pi K_3 - K_2) > 0$. It follows that $\pi K_3 - K_2 > 0$. ■

Proposition 1: The optimal work load for new associates, $\hat{n}$, for pairs (ρ^u, ρ^x) such that $0 \leq \rho^u + \rho^x < q$ exceeds the static optimum implied by the expected per period output of new associates.

Proof of Proposition 1: The first order condition for $\hat{n}$ when $0 \leq \rho^u + \rho^x < q$ is given by

$$c'(\hat{n}) = (1+\pi x) + \beta\phi'(\hat{n})(\pi K_3 - K_2) \tag{8.14}$$

From Lemma 1, $\pi K_3 - K_2 > 0$. Hence,

$$c'(\hat{n}) > 1 + \pi x$$

while the static optimum solves $c'(n) = 1 + \pi x$. Since c is strictly convex, it follows that $\hat{n}$ exceeds the static optimum. ■

Proposition 2: When $0 \leq \rho^u + \rho^x < q$, we have $\hat{\alpha}^u = 0$ if $v_1^a - v_0^a(\hat{n}) < \beta\pi\phi(\hat{n})(v_1^p - v_1^a)$

Proof of Proposition 2: The first-order condition for $\hat{\alpha}^u$ when $0 \leq \rho^u + \rho^x < q$ is given by

$$\hat{\alpha}^u = \begin{cases} 1 & \text{if } v_1^a - v_0^a - \beta\left[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2\right] > 0 \\ [0,1] & \text{if } v_1^a - v_0^a - \beta\left[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2\right] = 0 \\ 0 & \text{if } v_1^a - v_0^a - \beta\left[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2\right] < 0 \end{cases} \tag{8.15}$$

Substituting in for K_2 and K_3 from the proof of Claim 2, we have

$$\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2 = \frac{\pi\phi(\hat{n})(v_1^p - v_0^a) + \hat{\alpha}^u(1-\phi(\hat{n}))(v_1^a - v_0^a)}{1 + \beta\hat{\alpha}^u(1-\phi(\hat{n})) + \beta\pi\phi(\hat{n})} \tag{8.16}$$

and so $\hat{\alpha}^u = 1$ whenever

$$v_1^a - v_0^a \geq \beta\frac{\pi\phi(\hat{n})(v_1^p - v_0^a) + \hat{\alpha}^u(1-\phi(\hat{n}))(v_1^a - v_0^a)}{1 + \beta\hat{\alpha}^u(1-\phi(\hat{n})) + \beta\pi\phi(\hat{n})}$$

which implies

$$\begin{aligned}(1+\beta\pi\phi(\hat{n}))(v_1^a - v_0^a) &\geq \beta\pi\phi(\hat{n})(v_1^p - v_0^a) \\ (v_1^a - v_0^a) &\geq \beta\pi\phi(\hat{n})(v_1^p - v_1^a)\end{aligned} \tag{8.17}$$

The optimal $\hat{\alpha}^u$ is therefore given by

$$\hat{\alpha}^u = \begin{cases} 1 & \text{if } v_0^a(n) < v_1^a - \beta\pi\phi(\hat{n})(v_1^p - v_1^a) \\ [0,1] & \text{if } v_0^a(n) = v_1^a - \beta\pi\phi(\hat{n})(v_1^p - v_1^a) \\ 0 & \text{if } v_0^a(n) > v_1^a - \beta\pi\phi(\hat{n})(v_1^p - v_1^a) \end{cases} \tag{8.18}$$

which proves the result. ■

Proposition 3: The optimal work load $\hat{n}$ for new associates for pairs (ρ^u, ρ^x) such that $0 \leq \rho^u + \rho^x < q$ is increasing in z^p and weakly decreasing in z^a.

Proof of Proposition 3: Consider a constrained planner's problem in which α^u is given and the planner can only choose effort n, i.e.

$$V(\rho^u, \rho^x; \alpha^u) = \max_n s(n, \alpha^u, 1) + \beta V\left(\rho_{+1}^u, \rho_{+1}^x; \alpha^u\right)$$

where

$$\begin{aligned}\rho_{+1}^u &= (q - \rho^u\alpha^u - \rho^x)(1-\phi(n)) \\ \rho_{+1}^x &= (q - \rho^u\alpha^u - \rho^x)\pi\phi(n)\end{aligned}$$

Using the same argument as in the proof of Claim 1, we can show that $V(\rho^u, \rho^x; \alpha^u)$ is linear over the set of all points (ρ^u, ρ^x) for which $0 \leq \rho^u + \rho^x < q$, i.e. $V(\rho^u, \rho^x; \alpha^u) = K_1^* + K_2^*\rho^u + K_3^*\rho^x$ for some constants K_1^*, K_2^*, and K_3^*. This implies the optimal effort level n is independent of ρ^u and ρ^x over this region. Let $n^*(\alpha^u)$ denote the optimal effort level in this constrained problem. Note that $\hat{n}$, the effort level that solves the unconstrained planner's problem is equal to $n^*(\alpha^u)$ when $\alpha^u = \hat{\alpha}^u$, i.e. the constrained optimal effort level when the retention decision is set optimally.

Since the same $n^*(\alpha^u)$ maximizes the value function $V(\rho^u, \rho^x; \alpha^u)$ for all (ρ^u, ρ^x) for which $0 \leq \rho^u + \rho^x < q$, it must also maximize the value function at $(0,0)$. But $V(0,0;\alpha^u) = K_1^*$, where K_1^* is defined by the system of equations

$$\begin{aligned}K_1^* &= (2-q)v^o + qv_0^a + \beta q\left[\pi\phi(n^*(\alpha^u))K_3^* + (1-\phi(n^*(\alpha^u)))K_2^*\right] + \beta K_1^* \\ K_2^* &= \alpha^u(v_1^a - v_0^a) - \beta\alpha^u(\pi\phi(n^*(\alpha^u))K_3^* + (1-\phi(n^*(\alpha^u)))K_2^*) \\ K_3^* &= (v_1^p - v_0^a) - \beta(\pi\phi(n^*(\alpha^u))K_3^* + (1-\phi(n^*(\alpha^u)))K_2^*)\end{aligned}$$

Hence, $n^*(\alpha^u)$ must satisfy the first and second order necessary conditions

$$\frac{\partial K_1^*}{\partial n} = 0 \tag{8.19}$$

$$\frac{\partial^2 K_1^*}{\partial n^2} < 0 \tag{8.20}$$

To see how n^* varies with z^p, we can look at how it varies with v_1^p given the latter is monotonically increasing in z^p. Totally differentiate (8.19) to obtain

$$\frac{dn^*}{dv_1^p} = -\frac{\partial^2 K_1^*/\partial v_1^p \partial n}{\partial^2 K_1^*/\partial n^2}$$

Using the expressions for K_2^* and K_3^* from $V(\rho^u, \rho^x; \alpha^u)$ one can show[36] that

$$\frac{\partial^2 K_1^*}{\partial v_1^p \partial n} = \frac{q\beta\pi (1 + \alpha^u \beta) \phi'(n^*)}{(1 + \beta\alpha^u (1 - \phi(n^*)) + \beta\pi\phi(n^*))^2}$$

The expression for $\partial^2 K_1 / \partial n^2$ is given by

$$\frac{\partial^2 K_1^*}{\partial n^2} = \frac{q\zeta(n^*)}{(1 + \beta\alpha^u (1 - \phi(n^*)) + \beta\pi\phi(n^*))^2}$$

where

$$\begin{aligned} \zeta(n^*) &= -c''(n^*) (1 + \beta\alpha^u (1 - \phi(n^*)) + \beta\pi\phi(n^*)) + \\ & \beta\phi''(n^*) [\pi (v_1^p - v_0^a) - \alpha^u (v_1^a - v_0^a) + \pi\beta\alpha^u (v_1^p - v_1^a)] \end{aligned} \tag{8.21}$$

Since the necessary second-order condition implies $\frac{\partial^2 K_1^*}{\partial n^2} < 0$, we know that $\zeta < 0$. Taking the ratio of the two expressions reveals that

$$\frac{dn^*}{dv_1^p} = -\frac{\beta\pi (1 + \alpha^u \beta) \phi'(n^*)}{\zeta(n^*)} > 0$$

In other words, increasing v_1^p will induce the planner to choose a higher n^* for a given value of α^u.

By an analogous argument,

$$\frac{dn^*}{\partial v_1^a} = -\frac{\partial^2 K_1^* / \partial v_1^a \partial n}{\partial^2 K_1 / \partial n^2}$$

Using the expressions for K_2 and K_3 from $V(\rho^u, \rho^x; \alpha^u)$, we have

$$\frac{\partial^2 K_1^*}{\partial v_1^a \partial n} = -\frac{q\alpha^u \beta (1 + \pi\beta) \phi'(n^*)}{(1 + \beta\alpha^u (1 - \phi(n^*)) + \beta\pi\phi(n^*))^2}$$

and using the expression for $\frac{\partial^2 K_1^*}{\partial n^2}$ from above we have

$$\frac{dn^*}{dv_1^a} = \frac{\alpha^u \beta (1 + \pi\beta) \phi'(n^*)}{\zeta(n)} \leq 0.$$

This expression is strictly negative if $\alpha^u > 0$ and 0 otherwise.

Finally, we know from (8.18) that for any value of v_1^p, the optimal $\hat{\alpha}^u$ is either uniquely equal to 0, uniquely equal to 1, or else any value between 0 and 1 is optimal. When any value of α^u between 0 and 1 is optimal, the objective function in the unconstrained planner's problem is independent of α^u. Hence, the value of n we chose to maximize K_1 will be the same whether we set $\hat{\alpha}^u = 0$ or $\hat{\alpha}^u = 1$. This implies that the value of n^* that maximizes K_1^* when $\hat{\alpha}^u = 0$ is the same that maximizes K_1^* when $\hat{\alpha}^u = 1$, i.e. $n^*(0) = n^*(1)$ whenever all values in $[0,1]$ are optimal for α^u.

Although the optimal $\hat{\alpha}^u$ in (8.18) is a correspondence if v_1^p that can take on multiple values for some realizations of v_1^p, let us define a function $\widetilde{\alpha}^u(v_1^p)$ that is equal to the optimal $\hat{\alpha}^u$ for any value of v_1^p where the $\hat{\alpha}^u(v_1^p)$ is unique. For any v_1^p such that $\hat{\alpha}^u$ contains the entire interval $[0,1]$, we set $\widetilde{\alpha}^u(v_1^p) = 1$. Now, consider the function $n^*(\widetilde{\alpha}^u(v_1^p))$. This function assigns each value of v_1^p to the optimal value of n that solves the constrained planner's problem that chooses n

[36] We verify this using Mathematica. Code is available upon request.

for a given α^u, where the value of α^u corresponds to $\widetilde{\alpha}^u(v_1^p)$. As we already noted above, the unconstrained optimal effort level $\hat{n}$ is equal to $n^*(\hat{\alpha}^u)$. Hence, $\hat{n}(v_1^p) = n^*(\widetilde{\alpha}^u(v_1^p))$. That is, we can express the optimal effort level $\hat{n}$ for each v_1^p using the optimal effort from the constrained effort level $n^*(\alpha^u)$ by setting $\alpha^u = \widetilde{\alpha}^u(v_1^p)$. Our construction of $\widetilde{\alpha}^u(v_1^p)$ implies we can partition the set of all v_1^p into a set of closed intervals (that can include single points) over which $\widetilde{\alpha}^u(v_1^p) = 1$ and a set of open intervals over which $\widetilde{\alpha}^u(v_1^p) = 0$. The function $n^*(\widetilde{\alpha}^u(v_1^p))$ is increasing in each of these intervals, and is continuous for all v_1^p given our previous observation that $n^*(0) = n^*(1)$ whenever both 0 and 1 are optimal. This implies $\hat{n}$ is increasing in v_1^p. By a similar logic, we can define $\widetilde{\alpha}^u(v_1^a)$ to argue that $\hat{n}(v_1^p) = n^*(\widetilde{\alpha}^u(v_1^a))$, and use this argument to show that $\hat{n}(v_1^a)$ is non-increasing in v_1^a.

Proposition 4: The optimal $\hat{\alpha}^u$ for pairs (ρ^u, ρ^x) such that $0 \leq \rho^u + \rho^x < q$ is weakly decreasing in z^p and weakly increasing in z^a.

Proof of Proposition 4: Once again, we can conduct comparative statics with respect to v^p and v^a, since these are monotonically increasing in z^p and z^a, respectively. Let $\hat{\alpha}^u(v_1^p)$ denote the solution to the planner's problem as a function of v_1^p. Since the planner's objective function is continuous, $\hat{\alpha}^u(v_1^p)$ must be an upper hemicontinuous correspondence in v_1^p. This means that if there exists some value $\overline{v}$ such that $\lim_{v_1^p \to \overline{v}^-} \hat{\alpha}^u(v_1^p) \neq \lim_{v_1^p \to \overline{v}^+} \hat{\alpha}^u(v_1^p)$, then both the values $\lim_{v_1^p \to \overline{v}^-} \hat{\alpha}^u(v_1^p)$ and $\lim_{v_1^p \to \overline{v}^+} \hat{\alpha}^u(v_1^p)$ belong to the set of values $\hat{\alpha}^u(v_1^p)$ evaluated at $v_1^p = \overline{v}$. In other words, if the optimal value for α^u changes between 0 and 1 as we vary v_1^p, then at any value of v_1^p in which the optimal value of α^u switches between 0 and 1, the planner would be optimizing by setting α^u to either 0 or 1. Thus, at any such value, both 0 and 1 are optimal choices. In what follows, we will argue that there exists at most one value of v_1^p at which both 0 and 1 can simultaneously be optimal values for α^u. This implies that as we vary v_1^p, the optimal $\hat{\alpha}^u$ will change values at most once. The same logic applies to varying v_1^a holding all other parameters fixed. By appealing to boundary conditions, we can then say whether as we increase either v_1^p or v_1^a, the optimal $\hat{\alpha}^u$ must rise or fall in v_1^p and v_1^a, respectively.

In the proof of Proposition 3, we argued that whenever both $\alpha^u = 0$ and $\alpha^u = 1$ are optimal, we must have $n^*(0) = n^*(1)$. We now argue that all else fixed, there exists at most one value of v_1^p for which $n^*(0) = n^*(1)$. With only one such value for v_1^p, we can conclude that there exists at most one value of v_1^p for which both 0 and 1 are optimal values for α^u. A similar argument follows for v_1^a.

To establish this, recall from the proof of Proposition 3, that

$$\begin{aligned} \frac{dn^*}{dv_1^p} &= -\frac{\beta\pi(1+\alpha^u\beta)\phi'(n^*)}{\zeta(n^*)} \\ \frac{dn^*}{dv_1^a} &= \frac{\beta\alpha^u(1+\pi\beta)\phi'(n^*)}{\zeta(n^*)} \end{aligned}$$

That is, the optimal value of n^* holding α^u fixed varies in a particular way with v_1^p and v_1^a. With respect to v_1^a, it is immediate that there can be at most value of v_1^a for which $n^*(0) = n^*(1)$, since $n^*(0)$ does not vary with v_1^a while $n^*(1)$ is decreasing with v_1^a. Hence, the optimal $\hat{\alpha}^u$ includes both 0 and 1 at most once. In the case of v_1^p, note that from Proposition 2, whenever $n^*(0) = n^*(1)$, we must have

$$v_1^a - v_0^a = \beta\pi\phi(n)(v_1^p - v_1^a)$$

Substituting this into $\zeta(n)$ in (8.21) implies that

$$\zeta(n) = (1 + \beta\pi\phi(n) + \beta\alpha^u(1 - \phi(n)))\left[-c''(n) + \beta\pi\phi''(n)\left(v_1^p - v_1^a\right)\right]$$

and so for any value of v_1^p for which $n^*(0) = n^*(1)$, we have

$$\frac{dn^*}{dv_1^p} = \frac{\beta\pi(1 + \beta\alpha^u)\phi'(n)}{(1 + \beta\pi\phi(n) + \beta\alpha^u(1 - \phi(n)))\left[c''(n) - \beta\pi\phi''(n)\left(v_1^p - v_1^a\right)\right]}$$

Differentiating this with respect to α^u yields

$$\frac{\partial^2 n^*}{\partial\alpha^u\partial v_1^p} = \frac{\pi\beta^2(1 + \pi\beta)\phi(n)\phi'(n)}{(1 + \beta\alpha^u(1 - m) + \beta\pi m)^2\left[c''(n^*) - \pi\phi''(n)\left(v_1^p - v_1^a\right)\right]} > 0$$

Hence, whenever $n^*(1; v_1^p) = n^*(0; v_1^p)$, the derivative of $n^*(1; v_1^p) - n^*(0; v_1^p)$ with respect to v_1^p is positive. This implies there can be at most one value v_1^p for which $n^*(1; v_1^p) - n^*(0; v_1^p) = 0$. Hence, the optimal $\hat{\alpha}^u$ will change values between 0 and 1 at most once.

Since we know $\hat{\alpha}^u$ switches at most once, we need to determine whether as we increase v_1^a and v_1^p, if there is a switch, whether the switch will be from 0 to 1 or from 1 to 0. To do this, we only need to determine what happens at extreme cases, taking into account the restrictions we impose on parameters. On the one hand, we can always let $v_1^p \to \infty$, since we do impose any upper bound on z^p. Since a partner generates arbitrarily large amounts of surplus, it will eventually be optimal to set $\alpha^u = 0$ and focus on identifying people who can be promoted to partner. Hence, if there is a transition as v_1^p increases, it must be from $\alpha^u = 1$ to $\alpha^u = 0$. With regards to v_1^a, although we impose a restriction that $z^a < w_0 < z^a(1 + \pi x)$, the second inequality was only imposed because without it there is no reason to retain an uncertain worker, making retention trivial. However, if we drop the requirement that $z^a(1 + \pi x) > w_0$, the planner's problem would remain unchanged. Hence, we can take the limit as $z^a \to 1$ to obtain a boundary condition for $\hat{\alpha}^u$. In the limit as $z^a \to 1$, it is optimal to set $\alpha^u = 0$ and employ a young worker who has some option value than an experienced worker who does not. This is true regardless of whether $1 + \pi x > w_0$ or not. Here we use the fact that since $q < 1$, there will always be a young worker employed in the outside sector. It follows that as we increase α^u, if there is a transition, it must be from $\alpha^u = 0$ to $\alpha^u = 1$. ■

Appendix B: Learning About Learning

In this Appendix, we present a model of learning by doing given uncertain learning efficiency. We analyze the model and compare the results it produces to the results we derive for the screening model in Appendix A.

B1. Environment. We structure our model to replicate our screening model above as much as possible. Once again, time is infinite and workers live for two periods. There are also two sectors: professional and outside. The marginal product of tasks performed in the outside sector is a constant, $w^o > 1$, that does not vary with worker skill or experience.

In the professional sector, total output is determined by the number of tasks performed and by how well workers of different skill levels sort to different tasks. We assume there are three types of tasks that can be performed in the professional sector: tasks one, two, and three. Further, we assume that workers may possess one of three skill levels: high, medium, or low.

In this model, there are no shocks to output. Each type of task produces a constant marginal product as long as the worker assigned to the task meets the skill requirements for the task. If a worker attempts to perform a task that she is not qualified to perform, she produces negative output. For simplicity, we set this negative output level to $-\infty$.

Workers of all skill levels are qualified to perform task one, and this task yields a marginal product of one regardless of the skill level of the worker who performs it. Task two yields a marginal product $z^a > w^o > 1$ if the worker has a medium or high skill level, but low-skilled workers are unable to perform this task. Task three yields a marginal product of $z^p > z^a$ if the worker has a high skill level, but a worker at the medium or low skill level is not qualified to perform this task.

Each worker begins with a low skill level, which we define as a type l worker. If the worker enter the outside sector, she remains type l, but if the worker enters the professional sector, she learns by doing. She learns λ skills per task, n, she performs, where λ is a random variable drawn independently for each worker according to the distribution $F(\cdot)$, i.e. $\Pr(\lambda \leq x) = F(x)$. Workers know that their own λ values are drawn from $F(\cdot)$, but no individual worker has private information about her own learning efficiency, and no market participants possess private information about the learning efficiency of any new worker.

If a new associate with learning efficiency λ performs n tasks, she acquires s skills, where $s = \lambda n$. If s exceeds a cutoff level $\overline{s}$, the worker has reached the high skill level and is able to perform any of the three tasks in the professional sector. We refer to such a worker as a high type, or type h. If s is less than or equal to $\overline{s}$ but still greater than some lower cutoff $\underline{s} < \overline{s}$, the worker has achieved the medium skill level and is able to perform tasks one and two but not task three. This worker is a medium type, or type m. Finally, if s is less than or equal to $\underline{s}$, the worker remains at the low skill level, i.e. type l, and given our assumption, $w^o > 1$, her second period productivity is greatest in the outside sector.

Given our structure, the probabilities that a new associate who performs n tasks reaches the low, medium, or high skill levels are given by

$$\begin{aligned}
\phi_l(n) &\equiv \Pr(\lambda n \leq \underline{s}) = F\left(\frac{\underline{s}}{n}\right) \\
\phi_m(n) &\equiv \Pr(\lambda n \in (\underline{s}, \overline{s}]) = F\left(\frac{\overline{s}}{n}\right) - F\left(\frac{\underline{s}}{n}\right) \\
\phi_h(n) &\equiv \Pr(\lambda n > \overline{s}) = 1 - F\left(\frac{\overline{s}}{n}\right)
\end{aligned}$$

We maintain our assumptions about $c(n)$ from the screening model, and in particular that there is a maximal level of effort $\overline{n}$ workers can put in. Given our assumptions thus far, $\phi_h'(n) \geq 0$ and $\phi_l'(n) \leq 0$. Below, we impose a stronger condition that the derivative $\phi_h'(n)$ is strictly positive for n between the level of effort that maximizes first-period surplus, $c'^{-1}(1)$, and the maximal level of effort, $\overline{n}$. In the analyses below, we also assume that $\phi_h''(n) \leq 0$. Finally, we assume $\phi_l(\overline{n}) > 0$, so that even at the maximal effort level, there will be some workers who fail to learn enough skills to perform any tasks beyond task one. Note that any assumptions on the probabilities ϕ_j for $j \in \{l, m, h\}$ correspond to assumptions on the distribution of λ.

To create a structure that parallels our screening model, we refer to task three as the partner task, task two as the senior associate task, and task one as the new associate task. We define the surplus associated wtih these positions, v_1^p ,v_1^a, and $v_0^a(n)$, as before. That is,

$$\begin{aligned} v_0^a(n) &= n - c(n) \\ v_1^a &= \max_n z_m^a n - c(n) \\ v_1^p &= \max_n z_h^p n - c(n) \end{aligned}$$

Finally, the surplus on the outside sector is again defined as $v^o = \max_n w^o n - c(n)$.

Let ρ^h and ρ^m denote the fraction of experienced workers who are high and medium types, respectively. The planner's problem for our learning by doing model is given by

$$\begin{aligned} V\left(\rho^m, \rho^h\right) &= \max_{\alpha^h, \alpha^m, n} (2-q) v^o + q v_0^a(n) + \alpha^h \rho^h \left(v_1^p - v_0^a(n)\right) + \\ &\qquad \alpha^m \rho^m \left(v_1^a - v_0^a(n)\right) + \beta V\left(\rho_{+1}^m, \rho_{+1}^h\right) \end{aligned} \tag{8.22}$$

where

$$\rho_{+1}^m = \left(q - \rho^m \alpha^m - \rho^h \alpha^h\right) \phi_m(n) \tag{8.23}$$

$$\rho_{+1}^h = \left(q - \rho^m \alpha^m - \rho^h \alpha^h\right) \phi_h(n) \tag{8.24}$$

and the constraint

$$q - \alpha^m \rho^m - \alpha^h \rho^h \geq 0 \tag{8.25}$$

The state variables ρ^m and ρ^h are analogous to ρ^u and ρ^x in the original screening model. The only features that do not have exact parallels in our screening model are in the laws of motion (8.23) and (8.24). In particular, in the screening model, the analogs to $\phi_h(n)$ and $\phi_m(n)$ were required to satisfy $\phi_h(n) = \pi(1 - \phi_m(n))$, whereas now there is no analogous restriction. The counterpart to the restriction that $\phi'(n) > 0$ is that now $\phi_h'(n) > 0$. In the screening model, this would have implied $\phi_m'(n) < 0$, but this need not be the case in the present model, where the sign of $\phi_m'(n)$ is generally ambiguous. In what follows, we show that if $\phi_m'(n) \leq 0$, our learning by doing model yields results that parallel all of the results from our screening model. We also analyze the case where $\phi_m'(n) > 0$. In this case, we can quickly establish some results that parallel those from our screening model, but we need to impose an additional assumption to establish results that parallel all of findings in sections 3 and 4.

We begin by deriving results that are analogous to Propositions 1 and 2 in the screening model. We start here because we can establish these results without placing any restrictions on the sign of $\phi_m^{'}(n)$.

B2. Results that are Independent of the Sign of ϕ_m'. Applying the same arguments as in the proof of Claim 1, we can confirm that the Bellman equation for the planner's problem is once again linear in ρ^m and ρ^h whenever $0 \leq \rho^m + \rho^h < q$, i.e.

$$V\left(\rho^m, \rho^h\right) = K_1 + K_m \rho^m + K_h \rho^h$$

Using hats to denote optimal values, the coefficients K_1, K_m, and K_h satisfy the system of equations

$$\begin{aligned} K_1 &= (2-q)\,v^o + v_0^a(\hat{n}) + q\beta\left[\phi_m(\hat{n})\,K_m + \phi_h(\hat{n})\,K_h\right] + \beta K_1 \\ K_m &= \hat{\alpha}^m\left(v_1^a - v_0^a\right) - \beta\hat{\alpha}^m\left[\phi_m(\hat{n})\,K_m + \phi_h(\hat{n})\,K_h\right] \\ K_h &= \hat{\alpha}^h\left(v_1^p - v_0^a\right) - \beta\hat{\alpha}^h\left[\phi_m(\hat{n})\,K_m + \phi_h(\hat{n})\,K_h\right] \end{aligned}$$

Solving the above system yields

$$K_m = \frac{\hat{\alpha}^m\left(v_1^a - v_0^a\right) - \beta\hat{\alpha}^m\hat{\alpha}^h\phi_h(\hat{n})\left(v_1^p - v_1^a\right)}{1 + \beta\left[\hat{\alpha}^m\phi_m(\hat{n}) + \hat{\alpha}^h\phi_h(\hat{n})\right]} \tag{8.26}$$

$$K_h = \frac{\hat{\alpha}^h\left(v_1^p - v_0^a\right) + \beta\hat{\alpha}^h\hat{\alpha}^m\phi_m(\hat{n})\left(v_1^p - v_1^a\right)}{1 + \beta\left[\hat{\alpha}^m\phi_m(\hat{n}) + \hat{\alpha}^h\phi_h(\hat{n})\right]} \tag{8.27}$$

Note that the fact that $\phi_h'' \le 0$, which is analogous to our previous assumption that $\phi'' \le 0$, is no longer sufficient to ensure that the planner's problem is concave in n. For that, we now need

$$\phi_h''(n)\,K_h + \phi_m''(n)\,K_m - c''(n) < 0$$

The first order necessary condition for the optimal $\hat{n}$ is now given by

$$c'(\hat{n}) = 1 + \beta\left(\phi_h'(\hat{n})\,K_h + \phi_m'(\hat{n})\,K_m\right)$$

Since the value function is linear in α^m and α^h, the optimal choice for these control variables when $0 \le \rho^m + \rho^h < q$ is still given by

$$\hat{\alpha}^m = \begin{cases} 1 & \text{if } v_1^a - v_0^a - \beta\left[\phi_m(\hat{n})\,K_m + \phi_h(\hat{n})\,K_h\right] \ge 0 \\ 0 & \text{if } v_1^a - v_0^a - \beta\left[\phi_m(\hat{n})\,K_m + \phi_h(\hat{n})\,K_h\right] < 0 \end{cases}$$

$$\hat{\alpha}^h = \begin{cases} 1 & \text{if } v_1^p - v_0^a - \beta\left[\phi_m(\hat{n})\,K_m + \phi_h(\hat{n})\,K_h\right] \ge 0 \\ 0 & \text{if } v_1^p - v_0^a - \beta\left[\phi_m(\hat{n})\,K_m + \phi_h(\hat{n})\,K_h\right] < 0 \end{cases}$$

As in the previous case, the fact that employment in the professional sector is equal to q, there will be at most q inexperienced associates at any period t. Since some of them will fail to master enough skills to perform new jobs, it follows that $\rho_{t+1}^m + \rho_{t+1}^h < q$. Once again, then, constraint (8.25) will not bind beyond the initial period.

We can now establish the analog of Claim 2 that when $0 \le \rho^m + \rho^h < q$, it will be optimal to set $\hat{\alpha}^h = 1$. From the first order condition above, this requires

$$v_1^p - v_0^a \ge \beta\left[\phi_m(\hat{n})\,K_m + \phi_h(\hat{n})\,K_h\right]$$

It will suffice to show that $K_m \le v_1^p - v_0^a$ and $K_h \le v_1^p - v_0^a$. Observe that

$$\begin{aligned} v_1^p &\equiv \max_n z_h^p n - c(n) \\ &> \max_n z_m^a n - c(n) \\ &\equiv v_1^a \end{aligned}$$

Since $v_1^p - v_1^a > 0$, we have

$$\begin{aligned} K_m &= \frac{\hat{\alpha}^m \left(v_1^a - v_0^a\right) - \beta \hat{\alpha}^m \hat{\alpha}^h \phi_h\left(\hat{n}\right)\left(v_1^p - v_1^a\right)}{1 + \beta\left[\hat{\alpha}^m \phi_m\left(\hat{n}\right) + \hat{\alpha}^h \phi_h\left(\hat{n}\right)\right]} \\ &\leq \frac{\hat{\alpha}^m \left(v_1^a - v_0^a\right)}{1 + \beta\left[\hat{\alpha}^m \phi_m\left(\hat{n}\right) + \hat{\alpha}^h \phi_h\left(\hat{n}\right)\right]} \\ &\leq v_1^a - v_0^a \\ &\leq v_1^p - v_0^a \end{aligned}$$

Next, consider K_h. Observe that

$$\begin{aligned} v_1^a &\equiv \max_n z_m^a n - c\left(n\right) \\ &\geq \max_n n - c\left(n\right) \\ &\geq v_0^a \end{aligned}$$

This implies $v_1^p - v_0^a \geq v_1^p - v_1^a$, and so

$$\begin{aligned} K_h &= \frac{\hat{\alpha}^h \left(v_1^p - v_0^a\right) + \beta \hat{\alpha}^h \hat{\alpha}^m \phi_m\left(\hat{n}\right)\left(v_1^p - v_1^a\right)}{1 + \beta\left[\hat{\alpha}^m \phi_m\left(\hat{n}\right) + \hat{\alpha}^h \phi_h\left(\hat{n}\right)\right]} \\ &\leq \frac{1 + \beta \hat{\alpha}^m \phi_m\left(\hat{n}\right)}{1 + \beta\left[\hat{\alpha}^m \phi_m\left(\hat{n}\right) + \hat{\alpha}^h \phi_h\left(\hat{n}\right)\right]} \hat{\alpha}^h \left(v_1^p - v_0^a\right) \\ &\leq v_1^p - v_0^a \end{aligned}$$

It follows that $\hat{\alpha}^h = 1$ is optimal for $0 \leq \rho^m + \rho^h < q$.

Next, we establish the analog of Proposition 1. When $0 \leq \rho^m + \rho^h < q$, a necessary condition for the optimal $\hat{n}$ is

$$c'\left(\hat{n}\right) = 1 + \beta\left(\phi_m'\left(\hat{n}\right) K_m + \phi_h'\left(\hat{n}\right) K_h\right)$$

Establishing our result requires a condition that is the analog of Lemma 1. Specifically, we need to show that

$$\phi_m'\left(\hat{n}\right) K_m + \phi_h'\left(\hat{n}\right) K_h > 0$$

Substituting in $\hat{\alpha}^h = 1$ yields

$$\begin{aligned} K_m &= \frac{\hat{\alpha}^m \left(v_1^a - v_0^a\right) - \beta \hat{\alpha}^m \phi_h\left(\hat{n}\right)\left(v_1^p - v_1^a\right)}{1 + \beta\left[\hat{\alpha}^m \phi_m\left(\hat{n}\right) + \phi_h\left(\hat{n}\right)\right]} \\ K_h &= \frac{\left(v_1^p - v_0^a\right) + \beta \hat{\alpha}^m \phi_m\left(\hat{n}\right)\left(v_1^p - v_1^a\right)}{1 + \beta\left[\hat{\alpha}^m \phi_m\left(\hat{n}\right) + \phi_h\left(\hat{n}\right)\right]} \end{aligned}$$

Since $v_1^p > v_1^a$, then $K_h > K_m$. Hence,

$$\begin{aligned} \phi_m'\left(\hat{n}\right) K_m + \phi_h'\left(\hat{n}\right) K_h &> \phi_m'\left(\hat{n}\right) K_m + \phi_h'\left(\hat{n}\right) K_m \\ &= \left(\phi_m'\left(\hat{n}\right) + \phi_h'\left(\hat{n}\right)\right) K_m \\ &\geq 0 \end{aligned}$$

where the first inequality uses the fact that $\phi_h' > 0$ and the last inequality uses the fact that $\phi_m' + \phi_h' = -\phi_l' \geq 0$.

We can likewise establish the analog of Proposition 2. Setting $\hat{\alpha}^h = 1$, we get

$$\phi_m\left(\hat{n}\right) K_m + \phi_h\left(\hat{n}\right) K_h = \frac{\phi_h\left(\hat{n}\right)\left(v_1^p - v_0^a\right) + \hat{\alpha}^m \phi_m\left(\hat{n}\right)\left(v_1^a - v_0^a\right)}{1 + \beta\left(\phi_h\left(\hat{n}\right) + \hat{\alpha}^m \phi_m\left(\hat{n}\right)\right)}$$

Hence, $\hat{\alpha}^m = 1$ whenever

$$v_1^a - v_0^a \geq \beta \frac{\phi_h(\hat{n})(v_1^p - v_0^a) + \hat{\alpha}^m \phi_m(\hat{n})(v_1^a - v_0^a)}{1 + \beta(\phi_h(\hat{n}) + \alpha^m \phi_m(\hat{n}))}$$

which implies

$$\begin{aligned} (1 + \beta\phi_h(\hat{n}))(v_1^a - v_0^a) &\geq \beta\phi_h(\hat{n})(v_1^p - v_0^a) \\ v_1^a - v_0^a &\geq \beta\phi_h(\hat{n})(v_1^p - v_1^a) \end{aligned}$$

as desired.

B3. Results for the Case where $\phi_m' \leq 0$. We begin with the analog to Proposition 3. As we did in the screening model, let $n^*(\alpha^m)$ denote the level of effort which solves the planner's problem for a given value of α^m. It must satisfy the first order necessary condition

$$\frac{\partial K_1^*}{\partial n} = 0 \tag{8.28}$$

as well as the second order necessary condition

$$\frac{\partial^2 K_1^*}{\partial n^2} < 0 \tag{8.29}$$

Totally differentiating the first order condition yields

$$\frac{dn^*}{dv_1^p} = -\frac{\partial^2 K_1^* / \partial v_1^p \partial n}{\partial^2 K_1^* / \partial n^2}$$

With a little algebra (simplified via Mathematica), we have

$$\frac{\partial^2 K_1^*}{\partial v_1^p \partial n} = q\beta \frac{\phi_h' + \beta\alpha^m(\phi_h'\phi_m - \phi_m'\phi_h)}{(1 + \beta\phi_h + \beta\alpha^m\phi_m)^2}$$

When $\phi_m' \leq 0$, this expression is positive. Next, we have

$$\frac{\partial^2 K_1^*}{\partial n^2} = \frac{q\zeta(n)}{(1 + \beta\phi_h + \beta\alpha^m\phi_m)^2}$$

where

$$\begin{aligned} \zeta \equiv\ & \beta\phi_h''[v_1^p - v_0^a + \beta\alpha^m\phi_m(v_1^p - v_1^a)] + \beta\phi_m''[v_1^a - v_0^a - \beta\phi_h(v_1^p - v_1^a)] \\ & - c''(n)(1 + \beta\phi_h + \beta\alpha^m\phi_m) \end{aligned}$$

Since $(1 + \beta\phi_h + \beta\alpha^m\phi_m)^2 > 0$, the second order necessary condition for n to maximize K_1 implies $\zeta < 0$. Taking the ratio of the two expressions reveals that

$$\frac{dn^*}{dv_1^p} = -\beta \frac{\phi_h' + \beta\alpha^m[\phi_h'\phi_m - \phi_m'\phi_h]}{\zeta}$$

Since $\zeta < 0$ at the optimum, it follows from the second order condition that $\frac{dn^*}{dv_1^p} > 0$. Analogously, we have

$$\frac{\partial^2 K_1^*}{\partial v_1^a \partial n} = q\beta\alpha^m \frac{\phi_m' + \beta[\phi_m'\phi_h - \phi_h'\phi_m]}{(1 + \beta\phi_h + \beta\alpha^m\phi_m)^2}$$

If $\phi'_m \leq 0$, this derivative will be negative, in which case

$$\begin{aligned}\frac{dn^*}{dv_1^a} &= -\frac{\partial^2 K_1^* / \partial v_1^a \partial n}{\partial^2 K_1^* / \partial n^2} \\ &= -\beta\alpha^m \frac{\phi'_m + \beta\left[\phi'_m \phi_h - \phi'_h \phi_m\right]}{\zeta} \leq 0\end{aligned}$$

Again, as in the proof of Proposition 3, we can use the fact that $\hat{n}\left(v_1^p\right) = n^*(\hat{\alpha}^m)$ to establish that $\hat{n}\left(v_1^p\right)$ is a continuous piecewise increasing function and that $\hat{n}\left(v_1^a\right)$ is a continuous piecewise nonincreasing function. It follows that the Proposition 3 extends to the current setting.

To establish the analog of Proposition 4, we argue as before that there is at most one value of v_1^p and one value of v_1^a, respectively, for which $n^*(0) = n^*(1)$. In the case of v_1^a, this is once again immediate: $n^*(0)$ does not vary with v_1^a while $n^*(1)$ is decreasing with v_1^a, so they can equal at most once. In the case of v_1^p, once again we argue that whenever $n^*(1; v_1^p) - n^*(0; v_1^p) = 0$, the derivative of $n^*(1) - n^*(0)$ with respect to v_1^p is negative whenever $n^*(1) = n^*(0)$. Recall that $n^*(1) = n^*(0)$ iff

$$v_1^a - v_0^a = \beta\phi_h\left(v_1^p - v_1^a\right)$$

Hence, whenever $n^*(1) = n^*(0)$, the expression ζ reduces to

$$\zeta = \beta\left(1 + \beta\phi_h + \beta\alpha^m\phi_m\right)\left(\beta\phi''_h\left(v_1^p - v_1^a\right) - c''(n)\right)$$

and so

$$\frac{dn^*}{dv_1^p} = \frac{\beta\phi'_h + \beta^2\alpha^m\left[\phi'_h\phi_m - \phi'_m\phi_h\right]}{\left(c''(n) - \beta\phi''_h\left(v_1^p - v_1^a\right)\right)\left(1 + \beta\phi_h + \beta\alpha^m\phi_m\right)}$$

Differentiating with respect to α^m yields

$$\frac{\partial^2 n^*}{\partial\alpha^m \partial v_1^p} = \frac{\beta^2\phi_h\left(\beta\left(\phi'_h\phi_m - \phi'_m\phi_h\right) - \phi'_m\right)}{\left(c''(n) - \beta\phi''_h\left(v_1^p - v_1^a\right)\right)\left(1 + \beta\phi_h + \beta\alpha^m\phi_m\right)^2}$$

When $\phi''_h \leq 0$, the above expression must be positive when $\phi'_m < 0$, and the analysis follows as in the screening case.

B4. Results for the Case where $\phi'_m > 0$. We now turn to the case where $\phi'_m > 0$. The optimal assignment in this case depends on how the ratio ϕ_h/ϕ_m varies with n. Note that

$$\frac{d\left(\phi_h/\phi_m\right)}{dn} = \frac{\phi'_h\phi_m - \phi'_m\phi_h}{\left(\phi'_m\right)^2}$$

Whether ϕ_h/ϕ_m increases or decreases with n depends on the sign of the expression $\phi'_h\phi_m - \phi'_m\phi_h$. When $\phi'_m \leq 0$, as we considered in the previous section, the ratio ϕ_h/ϕ_m is strictly increasing in n.

As a benchmark, consider the case where ϕ_h/ϕ_m does not vary with n, i.e. where $\phi'_h\phi_m - \phi'_m\phi_h = 0$. In this case, we have

$$\begin{aligned}\frac{dn^*}{dv_1^p} &= -\frac{\beta\phi'_h}{\zeta} > 0 \\ \frac{dn^*}{dv_1^a} &= -\frac{\beta\alpha^m\phi'_m}{\zeta} \geq 0\end{aligned}$$

In this case, first-period effort is still strictly increasing in v_1^p, as in Proposition 3, but is weakly increasing in v_1^a, the opposite of what we found in Proposition 3.

As for Proposition 4, we can again show that $\hat{\alpha}^m$ will switch values at most once as we vary either v_1^p or v_1^a. As in Appendix A, this result is immediate for v_1^a. For v_1^p, the result follows because

$$\frac{\partial^2 n^*}{\partial \alpha^m \partial v_1^p} = \frac{-\beta^2 \phi_h \phi_m'}{\left(c''(n) - \beta \phi_h'' \left(v_1^p - v_1^a\right)\right)\left(1 + \beta \phi_h + \beta \alpha^m \phi_m\right)^2} < 0$$

as long as $\phi_h^{''} \leq 0$, which is the counterpart to the argument used in the proof of Proposition 4 above. Hence, $\hat{\alpha}^m$ is weakly monotonic in v_1^p, and the analog of Proposition 4 continues to hold.

Once we allow ϕ_h/ϕ_m to vary with n, the analysis becomes more complicated. Consider first the case where ϕ_h/ϕ_m is increasing in n for all n, meaning $\phi_h' \phi_m - \phi_m' \phi_h > 0$. In this case,

$$\frac{dn^*}{dv_1^p} = -\beta \frac{\phi_h' + \beta \alpha^m \left[\phi_h' \phi_m - \phi_m' \phi_h\right]}{\zeta} > 0$$

so $\dfrac{dn^*}{dv_1^p} > 0$ just as before. But the sign of $\dfrac{dn^*}{dv_1^a}$ is now ambiguous, since

$$\frac{dn^*}{dv_1^a} = -\beta \alpha^m \frac{\phi_m' - \beta \left[\phi_h' \phi_m - \phi_m' \phi_h\right]}{\zeta}$$

which given $\phi_m' > 0$ and $\phi_h' \phi_m - \phi_m' \phi_h > 0$ can be either positive or negative when $\hat{\alpha}^m = 1$. Intuitively, $\phi_m' > 0$ implies that as z^a rises, the direct returns to effort from new associates also rises; putting in more effort increases the odds of becoming a worker just skilled enough to perform the associate task, and that task becomes more productive when z^a rises. However, the gain associated with becoming a high type h rather than a medium type m is decreasing in z^a, and if increasing n shifts the relative odds for workers towards becoming high types rather than medium types, a worker may less have incentive to put in effort as z^a rises, since a higher n means he is less likely to end up with medium skill if he remains in the professional sector.

Although the sign of $\frac{dn^*(1)}{dv_1^a}$ is ambiguous, as long as ϕ_h' and ϕ_m' are continuous – which will be true if the distribution of learning abilities $F(\cdot)$ has no mass points – the sign of the derivative $\frac{dn^*(1)}{dv_1^a}$ will not switch as we vary v_1^a. This is because by continuity, the sign of this derivative can only switch if there exists a value for n^* at which $\phi_m' - \beta \left[\phi_h' \phi_m - \phi_m' \phi_h\right] = 0$. However, at any such point, $\frac{dn^*}{dv_1^a} = 0$. Since v_1^a only affects $\frac{dn^*}{dv_1^a} = 0$ through n^*, the existence of a value of v_1^a at which $\frac{dn^*}{dv_1^a} = 0$ implies that $\frac{dn^*}{dv_1^a} = 0$ for all values of v_1^a. This implies that $\phi_m' - \beta \left[\phi_h' \phi_m - \phi_m' \phi_h\right]$ cannot change signs as we vary v_1^a.

Since we know that $\frac{dn^*}{dv_1^a}$ will have the same sign for all v_1^a, we can conclude that if the ratio ϕ_h/ϕ_m increased sufficiently with n, specifically if $\phi_h' \phi_m - \phi_m' \phi_h > \phi_m'/\beta$, then Proposition 3 would continue to hold. If the ratio ϕ_h/ϕ_m were instead only modestly increasing in n, then just as in the case where ϕ_h/ϕ_m is invariant to n, first-period effort would still strictly be increasing in v_1^p as in Proposition 3, but would be weakly increasing in v_1^a, in contrast to Proposition 3.

As for the analog of Proposition 4, the result that $\hat{\alpha}^m$ can switch at most once as we vary v_1^a continues to hold, since $\frac{dn^*(0)}{dv_1^a} = 0$ while $\frac{dn^*(1)}{dv_1^a}$ is monotonic, although we cannot say whether $n^*(1)$ is increasing or decreasing in v_1^a. It follows that $\hat{\alpha}^m$ is weakly increasing in z^a, as in Proposition 4. With regards to how $\hat{\alpha}^m$ varies with z^p, note that the sign of $\dfrac{\partial^2 n^*}{\partial \alpha^m \partial v_1^p}$ is equal to the sign of $\beta\left[\phi_h' \phi_m - \phi_m' \phi_h\right] - \phi_m'$. If this sign were either positive for all z^p or negative for all z^p, then it would follow that $\hat{\alpha}^m$ can switch at most once as we vary v_1^p. Hence, if the ratio ϕ_h/ϕ_m increased sufficiently with n, the result of Proposition 4 would hold. Otherwise, we could not rule out the possibility that $\hat{\alpha}^m$ is non-monotonic in z^p.

Finally, consider the case where ϕ_h/ϕ_m is decreasing in n, i.e. $\phi_h'\phi_m - \phi_m'\phi_h < 0$. In this case,

$$\frac{dn^*}{dv_1^p} = -\beta \frac{\phi_h' + \beta\alpha^m\left[\phi_h'\phi_m - \phi_m'\phi_h\right]}{\zeta}$$

which can be either positive or negative, while

$$\frac{dn^*}{dv_1^a} = -\beta\alpha^m \frac{\phi_m' - \beta\left[\phi_h'\phi_m - \phi_m'\phi_h\right]}{\zeta} \geq 0$$

Intuitively, an increase in v_1^p would tend to lead to more hours in order to train more workers to be partners. However, if this disproportionately increases the fraction of middle types, it might be preferable to cut back on training, increase surplus, and leave more slots for identifying talent.

Since $\phi_h' > 0$, we know $\frac{dn^*(0)}{dv_1^p} > 0$. By the same logic as before, we know that as long as ϕ_h' and ϕ_m' are continuous, the sign of $\frac{dn^*(1)}{dv_1^p}$ will not switch as we vary v_1^p. If $\frac{dn^*(1)}{dv_1^p} \geq 0$, meaning that ϕ_h/ϕ_m was only modestly decreasing in n, then $\hat{n}\left(v_1^p\right)$ would be a continuous piecewise increasing function, confirming the first part of Proposition 3, while the second part of Proposition 3 would flip since $\hat{n}$ would be weakly increasing in v_1^a. Otherwise, even though the sign of $\frac{dn^*(1)}{dv_1^p}$ will not switch as we vary v_1^p, the function $\hat{n}\left(v_1^p\right)$ would be non-monotonic, so neither parts of Proposition 3 would hold.

As for the analog of Proposition 4, it will no longer be the case that $\hat{\alpha}^m$ must be weakly increasing in z^a as in Proposition 4, since $n^*(1)$ can be non-monotonic in v_1^p even if $n^*(0)$ is constant. However, since the sign of $\dfrac{\partial^2 n^*}{\partial \alpha^m \partial v_1^p}$ is equal to the sign of $\beta\left[\phi_h'\phi_m - \phi_m'\phi_h\right] - \phi_m'$, which we know is negative, we can establish that $\hat{\alpha}^m$ must be weakly decreasing in z^p.

Appendix C: Endogenous Sector Size and Decentralization

In this Appendix, we consider the case where the size of the sector q is endogenous. We first solve the planner's problem, and then show that it can be achieved as the decentralized equilibrium of a market economy.

C1. Planner's Problem with Endogenous Sector Size. When the number of jobs q is endogenous, the planner's problem is given by

$$V\left(\rho^u, \rho^x\right) = \max_{q, \alpha^x, \alpha^u, n} s\left(q, \alpha^x, \alpha^u, n\right) - \int_0^q \kappa(x)\, dx + \beta V\left(\rho_{+1}^u, \rho_{+1}^x\right)$$

subject to

$$\begin{aligned}\rho^u_{+1} &= (q-\rho^u\alpha^u-\rho^x\alpha^x)(1-\phi(n))\\ \rho^x_{+1} &= (q-\rho^u\alpha^u-\rho^x\alpha^x)\pi\phi(n)\end{aligned}$$

as well as the constraint

$$q-\rho^u_t\alpha^u-\rho^x_t\alpha^x\geq 0 \tag{8.30}$$

The above problem is similar to our original planning problem, except that rather than treating q as given, the planner now chooses q in addition to α^x, α^u, and n^u_0, and incurs a cost $\int_0^q \kappa(x)\,dx$ that depends on the number of jobs created in the professional sector.

We begin by ignoring (8.30) and analyzing an unconstrained planning problem that does not involve this constraint. We can then verify whether the solution to the unconstrained problem satisfies this constraint. Let ξ^x_0 denote the multiplier on the constraint that $\alpha^x\geq 0$ and ξ^x_1 denote the multiplier on the constraint that $\alpha^x\leq 1$. Likewise, let ξ^u_0 denote the multiplier on the constraint that $\alpha^u\geq 0$ and ξ^u_1 denote the multiplier on the constraint that $\alpha^u\leq 1$. The first order conditions of the maximization problem above with respect to n, α^x, α^u, and q are given by

$$n \quad : \quad (1+\pi x)-c'(n)-\beta\phi'(n)\left[\pi\frac{\partial V}{\partial\rho^x}-\frac{\partial V}{\partial\rho^u}\right]=0 \tag{8.31}$$

$$\alpha^x \quad : \quad \rho^x\left[(v^p_1-v^a_0)-\beta\left(\pi\phi(n)\frac{\partial V}{\partial\rho^x}-(1-\phi(n))\frac{\partial V}{\partial\rho^u}\right)\right]=\xi^x_1-\xi^x_0 \tag{8.32}$$

$$\alpha^u \quad : \quad \rho^u\left[(v^a_1-v^a_0)-\beta\left(\pi\phi(n)\frac{\partial V}{\partial\rho^x}-(1-\phi(n))\frac{\partial V}{\partial\rho^u}\right)\right]=\xi^u_1-\xi^u_0 \tag{8.33}$$

$$q \quad : \quad -v^o+v^a_0-\kappa(q)+\beta\left[\pi\phi(n)\frac{\partial V}{\partial\rho^x}+(1-\phi(n))\frac{\partial V}{\partial\rho^u}\right]=0 \tag{8.34}$$

Define $\mathbb{U}$ as the set of values (ρ^u,ρ^x) for which the optimal solution to the unconstrained problem implies $\hat{q}-\rho^u_t\hat{\alpha}^u-\rho^x_t\hat{\alpha}^x>0$. At such values, (8.30) will not be binding, so within this set the optimal plan to the constrained planner's problem coincides with the unconstrained problem.

As in the case where q was fixed, we first argue that the function V is linear in (ρ^u,ρ^x) over the set $\mathbb{U}$. This is because the operator T implied by the Bellman equation above maps functions V that are linear over the set $\mathbb{U}$ into functions $T(V)$ that are linear over the set $\mathbb{U}$. To see this, let us once again denote the optimal values with a hat. Equation (8.31) implies that the optimal $\hat{n}$ will not vary with (ρ^u,ρ^x) in the set $\mathbb{U}$ given V is linear over this set, meaning $\partial V/\partial\rho^x_t$ and $\partial V/\partial\rho^u_t$ are constant. Note here that the argument relies on the optimal solution satisfying the condition that $\hat{q}-\rho^u_t\hat{\alpha}^u-\rho^x_t\hat{\alpha}^x>0$. From (8.32) and (8.33), we can then conclude that the optimal $\hat{\alpha}^u$ and $\hat{\alpha}^x$ do not vary with (ρ^u,ρ^x) in the set $\mathbb{U}$. In particular, when $V(\rho^u,\rho^x)$ is linear and the derivatives in (8.32) and (8.33) are independent of (ρ^u,ρ^x), the derivative with respect to α^x and α^u will either be positive, zero, or negative for all $(\rho^u,\rho^x)\in\mathbb{U}$, in which case the optimal $\widehat{\alpha}^u$ and $\widehat{\alpha}^x$ will be the same, regardless of the value of the multipliers ξ^x_1, ξ^u_1, ξ^x_0, and ξ^u_0. Finally, from (8.34), we can say that the optimal $\hat{q}$ does not vary with (ρ^u,ρ^x) in the set $\mathbb{U}$. Given the optimal vector $(\hat{\alpha}^x,\hat{\alpha}^u,\hat{n},\hat{q})$ is constant over $\mathbb{U}$, the function $T(V)$ must be linear in (ρ^u,ρ^x) over the set $\mathbb{U}$ as well. Since the Bellman equation

V is the unique fixed point which solves $V = T(V)$, it follows that the Bellman equation is linear in (ρ^u, ρ^x) over the set $\mathbb{U}$, i.e.

$$V(\rho_t^u, \rho_t^x) = K_1 + K_2\rho_t^u + K_3\rho_t^x$$

for $(\rho_t^u, \rho_t^x) \in \mathbb{U}$. Solving for K_2 and K_3 in the same way as we did when we assumed q was exogenous shows these values are unchanged, and it is only the expression for K_1 that depends on the optimal choice $\hat{q}$.

Since the optimal $\hat{q}$ does not vary with (ρ^u, ρ^x) in the set $\mathbb{U}$, we can denote this common value by q^*. From (8.34), we know that q^* solves

$$\kappa(q^*) = \beta[\pi\phi(n)K_3 + (1-\phi(n))K_2] - v^o + v_0^a$$

Recall that K_2 and K_3 are defined independently of q, so the above yields a closed-form expression for q^*. We can now argue that the set $\mathbb{U}$ is nonempty. Consider the set $\Omega \equiv \left\{(\rho^u, \rho^x) \in \mathbb{R}_+^2 : \rho^u + \rho^x < q^*\right\}$. Since $q^* > 0$, the set Ω is nonempty. Moreover, since α^x and α^u are at most 1, it follows that $\Omega \subset \mathbb{U}$, ensuring the latter is nonempty. That is, ignoring constraint (8.30) yields a unique optimal sector size q^* the planner would prefer, and as long as the total number of experienced workers is below q^*, constraint (8.30) will not bind.

Finally, we argue that under the optimal plan, $(\rho_t^u, \rho_t^x) \in \mathbb{U}$ for all $t \geq 1$, i.e. constraint (8.30) will cease to bind after one period. At date 0, we can have $(\rho_0^u, \rho_0^x) \in \mathbb{U}$ or $(\rho_0^u, \rho_0^x) \notin \mathbb{U}$. If $(\rho_0^u, \rho_0^x) \in \mathbb{U}$, then the planner will choose $q = q^*$. In this case, we know from the law of motion that

$$\rho_{t+1}^u + \rho_{t+1}^x = (q^* - \rho_t^u\alpha^u - \rho_t^x\alpha^x)(1-(1-\pi)\phi(n))$$

For $t = 0$, the first term is at most q^*, while the second term is strictly less than 1 since the optimal n when $(\rho_0^u, \rho_0^x) \in \mathbb{U}$ is positive. Hence,

$$\rho_{t+1}^u + \rho_{t+1}^x < q^*$$

Since $\hat{\alpha}^x$ and $\hat{\alpha}^u$ are both between 0 and 1, this implies

$$\hat{\alpha}^u\rho_{t+1}^u + \hat{\alpha}^x\rho_{t+1}^x \leq \rho_{t+1}^u + \rho_{t+1}^x < q^*$$

That is, if $(\rho_t^u, \rho_t^x) \in \mathbb{U}$ then $\left(\rho_{t+1}^u, \rho_{t+1}^x\right) \in \mathbb{U}$. This leaves us with the case $(\rho_0^u, \rho_0^x) \notin \mathbb{U}$. In this case, (8.30) is strictly binding, meaning

$$q_0 - \rho_0^u\alpha_0^u - \rho_0^x\alpha_0^x = 0$$

It then follows that $\rho_1^u = \rho_1^x = 0$, in which case

$$\rho_1^u + \rho_1^x = 0 < q^*$$

i.e. $(\rho_1^u, \rho_1^x) \in \mathbb{U}$. But since we just argued that $(\rho_t^u, \rho_t^x) \in \mathbb{U}$ implies $\left(\rho_{t+1}^u, \rho_{t+1}^x\right) \in \mathbb{U}$, this means that the optimal path will ensure $(\rho_t^u, \rho_t^x) \in \mathbb{U}$ for all $t \geq 1$.

Finally, we establish the following comparative static result:

Proposition 5: The optimal size of the professional sector $\hat{q}$ for dates $t \geq 1$ is increasing in z^p and weakly increasing in z^a.

Proof of Proposition 5: From the first order condition for the planner's problem, we have

$$\begin{aligned} \kappa(\hat{q}) &= v_0^a - v^o + \beta\left[\pi\phi(\hat{n})\frac{\partial V}{\partial \rho^x} + (1-\phi(\hat{n}))\frac{\partial V}{\partial \rho^u}\right] \\ &= v_0^a - v^o + \beta[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2] \qquad (8.35) \end{aligned}$$

From Proposition 4, we know that $\hat{\alpha}^u$ is nondecreasing in v_1^p, and there exists a single value of v_1^p for which all values $[0,1]$ are optimal. Hence, without loss of generality, we can treat the optimal $\hat{\alpha}^u$ as fixed when we increase v_1^p by a small amount. If we fix $\hat{n}$, the expression

$$\pi\phi(\hat{n}) K_3 + (1-\phi(\hat{n})) K_2 = \frac{\pi\phi(\hat{n})(v_1^p - v_0^a) + \hat{\alpha}^u (1-\phi(\hat{n}))(v_1^a - v_0^a)}{1+\beta\hat{\alpha}^u(1-\phi(\hat{n})) + \beta\pi\phi(\hat{n})}$$

is strictly increasing in v_1^p and weakly increasing in v_1^a (and strictly increasing in v_1^a if $\hat{\alpha}^u = 1$). But since $\hat{n}$ must maximize K_1^* evaluated at the optimal $\hat{\alpha}^u$, it follows that

$$\frac{d}{dn}\left[v_0^a(\hat{n}) + \beta(\pi\phi(\hat{n}) K_3 + \beta(1-\phi(\hat{n})) K_2)\right] = 0$$

where we have replaced K_2^* and K_3^* with K_2 and K_3, the value for K_2^* and K_3^* when we set α^u to the optimal value, $\hat{\alpha}^u$. Hence, at the optimum, changing $\hat{n}$ will have no effect on the RHS of (8.35). Thus, at the optimal allocation, the RHS of (8.35) must be increasing in v_1^p and weakly increasing in v_1^a. Since $\kappa(q)$ is assumed to be increasing in q, the optimal $\hat{q}$ will be higher. ■

Decentralization. We now consider a market economy in which workers choose where to work and how many tasks, n, to perform. Those who work in the professional sector must hire resources as input, e.g. office space, tech support, paralegals, etc. For simplicity, we will refer to these inputs as support staff, and assume these workers are separate from the mass 2 of workers who choose between the professional and outside sector. Denote the price of support staff by p. An equilibrium is a rule that dictates what each worker type (i.e. past work experience and information about ability θ) chooses as her job and how much work each worker performs in each job, such that, given a price of support staff $\widetilde{p}$, worker choices produce a quantity $\widetilde{q}$ of professional workers (and support staff) such that (1) the sector and work load choices of all workers are optimal, and (2) the market for support staff clears.

Define $\widetilde{\alpha}^x$ as a variable equal to 1 if an experienced worker known to be the high type chooses to work as a professional in equilibrium, and 0 if such a worker chooses to work in the outside sector. Likewise, define $\widetilde{\alpha}^u$ as a variable equal to 1 if an experienced worker whose θ is unknown chooses to work as a professional and 0 if the worker chooses to work in the outside sector. Finally, let $\widetilde{n}$ denote the effort choice of an inexperienced worker who works in the professional sector. We want to confirm that the planner's optimal allocation $(\hat{q}, \hat{\alpha}^x, \hat{\alpha}^u, \hat{n})$ constitutes an equilibrium, together with the remaining occupation and effort choices in Table 1.

First, if $\hat{q}$ people are employed in the professional sector, $\hat{q}$ support staff must be hired, and the equilibrium price $\widetilde{p}$ must equal $\kappa(\hat{q})$, since this is the only price at which the supply of support staff is equal to $\hat{q}$. A worker who chooses to work in the professional sector in the current period will thus earn

$$\max_{n,j} E\left[z_i^j\right] n - c(n) - \kappa(\hat{q})$$

where z_i^j denotes the productivity in job j of worker of type i.

It is easy to verify that some of the choices workers will make are identical to what the planner chooses in Table 1. For example, experienced workers known to be the high type will work as partners if they stay in the professional sector while experienced workers whose type is unknown will work as associates if they stay in the professional sector. Experienced workers known to be low types will prefer to work in the outside sector where they are more productive. Experienced workers who choose to work in the professional sector choose the static level of effort which solves $c'(n) = \max_j E\left[z_i^j\right]$, and experienced workers in the outside sector will choose the effort level that solves

$$n^o = \arg\max_n w^o n - c(n)$$

Hence, the utility of an experienced worker known to be the high type who opts to work in the professional sector is $v_1^p - \kappa(\hat{q})$, and the utility of an experienced worker of unknown ability who opts to work in the professional sector is $v_1^a - \kappa(\hat{q})$.

We now verify that the optimal allocation $(\hat{q}, \hat{\alpha}^x, \hat{\alpha}^u, \hat{n})$ is indeed an equilibrium, and that workers who work in the outside sector when young prefer to work in the outside sector when old.

We first verify that experienced workers who know they are the high type prefer to work as partners than work in the outside sector, i.e. $\widetilde{\alpha}^x = 1$. This requires

$$v_1^p - \kappa(\hat{q}) \geq \max_n w^o n - c(n)$$

But from the first-order condition (8.34), we have

$$v^o = v_0^a + \beta\left[\pi\phi(n) K_3 + (1 - \phi(n)) K_2\right] - \kappa(\hat{q})$$

Substituting in for K_2 and K_3 from the planner's problem reveals that

$$\begin{aligned} v^o &= \frac{v_0^a + \beta\pi\phi(\hat{n}) v_1^p + \beta\hat{\alpha}^u (1 - \phi(\hat{n})) v_1^a}{1 + \beta\pi\phi(\hat{n}) + \beta\hat{\alpha}^u (1 - \phi(\hat{n}))} - \kappa(\hat{q}) \\ &< v_1^p - \kappa(\hat{q}) \end{aligned}$$

where last inequality uses the fact that $v_1^p > v_0^a$ and $v_1^p > v_1^a$. It follows that $v_1^p - \kappa(\hat{q}) > v^o$ and so if $\widetilde{q} = \hat{q}$, then $\widetilde{\alpha}^x = 1 = \hat{\alpha}^x$.

Next, we verify that experienced workers whose type is uncertain work in the professional sector iff the planner would assign such workers to the professional sector, i.e. iff $\hat{\alpha}^u = 1$. Again, using (8.34), we know that

$$v^o = \frac{v_0^a + \beta\pi\phi(\hat{n}) v_1^p + \beta\hat{\alpha}^u (1 - \phi(\hat{n})) v_1^a}{1 + \beta\pi\phi(\hat{n}) + \beta\hat{\alpha}^u (1 - \phi(\hat{n}))} - \kappa(\hat{q})$$

If it is optimal to let such workers go, meaning $\hat{\alpha}^u = 0$ is optimal, this expression would reduce to

$$v^o = \frac{v_0^a + \beta\pi\phi(\hat{n}) v_1^p}{1 + \beta\pi\phi(\hat{n})} - \kappa(\hat{q})$$

Moreover, we know from Proposition 2 that $\hat{\alpha}^u = 0$ iff

$$v_1^a \leq v_0^a + \beta\pi\phi(\hat{n})(v_1^p - v_1^a)$$

Rearranging this inequality implies

$$v_1^a \leq \frac{v_0^a + \beta\pi\phi(\hat{n}) v_1^p}{1 + \beta\pi\phi(\hat{n})}$$

and so

$$v_1^a - \kappa(\hat{q}) \leq \frac{v_0^a + \beta\pi\phi(\hat{n}) v_1^p}{1 + \beta\pi\phi(\hat{n})} - \kappa(\hat{q}) = v^o$$

Conversely, if $\hat{\alpha}^u = 1$ is optimal, then $v_1^a \geq v_0^a + \beta\pi(\hat{n})(v_1^p - v_1^a)$. This implies

$$(1 + \beta\pi(\hat{n}))\, v_1^a \geq v_0^a + \beta\pi(\hat{n})\, v_1^p$$

Substituting this into our expression for v^o when $\hat{\alpha}^u = 1$ implies

$$\begin{aligned} v^o &= \frac{v_0^a + \beta\pi\phi(\hat{n})\, v_1^p + \beta\alpha^u (1 - \phi(\hat{n}))\, v_1^a}{1 + \beta\pi\phi(\hat{n}) + \beta\alpha^u (1 - \phi(\hat{n}))} - \kappa(\hat{q}) \\ &\leq \frac{(1 + \beta\pi\phi(\hat{n}))\, v_1^a + \beta\alpha^u (1 - \phi(\hat{n}))\, v_1^a}{1 + \beta\pi\phi(\hat{n}) + \beta\alpha^u (1 - \phi(\hat{n}))} - \kappa(\hat{q}) \\ &= v_1^a - \kappa(\hat{q}) \end{aligned}$$

Thus, if $\widetilde{q} = \hat{q}$, then $\widetilde{\alpha}^u = 1 = \hat{\alpha}^u$ and experienced workers with uncertain ability work as associates in the professional sector only when the planner assigns them to work as associates.

Next, a new worker who starts in the professional sector will choose $\widetilde{n}$ to maximize

$$\begin{aligned} &v_0^a(\widetilde{n}) - \kappa(\hat{q}) + \beta\pi\phi(\widetilde{n})(v_1^p - \kappa(\hat{q})) + \beta(1-\pi)\phi(\widetilde{n})\, v^o + \\ &\qquad \beta(1 - \phi(\widetilde{n}))(1 - \alpha^u)\, v^o + \beta(1 - \phi(\widetilde{n}))\, \widetilde{\alpha}^u (v_1^a - \kappa(\hat{q})) \end{aligned}$$

This is a well-defined concave problem with first order condition given by

$$c'(\widetilde{n}) = 1 + \pi x + \beta\phi'(\widetilde{n})\left[\pi v_1^p - \widetilde{\alpha}^u v_1^a + (\widetilde{\alpha}^u - \pi)(v^o + \kappa(\hat{q}))\right] \tag{8.36}$$

From the planner's first order condition (8.34), substituting in for K_2 and K_3 and the fact that $\widetilde{\alpha}^u = \hat{\alpha}^u$, we can rewrite (8.36) as

$$\begin{aligned} c'(\widetilde{n}) &= 1 + \pi x + \beta\phi'(\widetilde{n})\left[\pi v_1^p - \hat{\alpha}^u v_1^a + (\hat{\alpha}^u - \pi)\frac{v_0^a + \beta\pi\phi(\hat{n})\, v_1^p + \beta\hat{\alpha}^u (1 - \phi(\hat{n}))\, v_1^a}{1 + \beta\pi\phi(\hat{n}) + \beta\hat{\alpha}^u (1 - \phi(\hat{n}))}\right] \\ &= 1 + \pi x + \beta\phi'(\widetilde{n})\left[\frac{\pi(1 + \beta\hat{\alpha}^u)\, v_1^p - \hat{\alpha}^u (1 + \pi\beta)\, v_1^a + (\hat{\alpha}^u - \pi)\, v_0^a}{1 + \beta\pi\phi(\hat{n}) + \beta\hat{\alpha}^u (1 - \phi(\hat{n}))}\right] \\ &= 1 + \pi x + \beta\phi'(\widetilde{n})\left[\pi K_3 - K_2\right] \end{aligned}$$

which confirms that $\widetilde{n} = \hat{n}$, since $\hat{n}$ is the unique solution to $c'(n) = 1 + \pi x + \beta\phi'(n)[\pi K_3 - K_2]$ for given constants K_2 and K_3.

Next, we verify that new workers are indifferent between the two sectors when $\widetilde{q} = \hat{q}$. This indifference is required because new workers must enter both the professional and the outside sector. Since young workers will choose to put in the optimal level of effort $\hat{n}$, indifference requires that

$$\begin{aligned} (1 + \beta)\, v^o &= v_0^a - \kappa(\hat{q}) + \beta\pi\phi(\hat{n})(v_1^p - \kappa(\hat{q})) + \beta(1-\pi)\phi(\hat{n})\, v^o + \\ &\quad\ \beta(1 - \phi(\hat{n}))(1 - \alpha^u)\, v^o + \beta(1 - \phi(\hat{n}))\, \alpha^u (v_1^a - \kappa(\hat{q})) \end{aligned}$$

which upon rearranging implies

$$v^o = \frac{v_0^a + \beta\pi\phi(\hat{n})\, v_1^p + \beta\alpha^u (1 - \phi(\hat{n}))\, v_1^a}{1 + \beta\pi\phi(\hat{n}) + \beta\alpha^u (1 - \phi(\hat{n}))} - \kappa(\hat{q})$$

but this is precisely the first-order condition for the optimal $\hat{q}$. Hence, with $\hat{q}$ workers in the professional sector, young workers with no experience will be indifferent between going to the professional sector and the outside sector.

Finally, since young workers are just indifferent between the two sectors, older workers who worked in the outside sector when young will strictly prefer to work in the outside sector. That is, we can rewrite the indifference condition for young workers as

$$v^o = v_0^a - \kappa\left(\hat{q}\right) + \beta\pi\phi\left(\hat{n}\right)\left(v_1^p - v^o - \kappa\left(\hat{q}\right)\right) + \beta\left(1-\phi\left(\hat{n}\right)\right)\alpha^u\left(v_1^a - v^o - \kappa\left(\hat{q}\right)\right)$$

Since $\hat{n}$ is optimal, we know that the RHS above is higher at $\hat{n}$ than at the static optimum, and the value at the static optimum, which in turn exceeds the value of working in the professional sector at the static optimal level for only one period. This further implies that workers who start in the professional sector are strictly worse off in their first period than workers who start in the outside sector, a result we refer to in the text.

Appendix D - Data

This appendix describes the data sets that we employ in section 7 of the paper. We describe two data sources and our procedures for selecting and cleaning the samples we draw from these sources.

Survey of Law Firm Economics. Table 2 uses data from the *Survey of Law Firm Economics* (SLFE). This survey is conducted annually by ALM Legal Intelligence. We obtained electronic versions of the data for the eight surveys conducted between 2007 and 2014.

ALM generates its sample from directories of law firms, and it relies heavily on its own client lists from previous years. Law firms purchase data from ALM to help them benchmark their performance against other law firms.

We received two data sets from ALM. One contains records that describe individual lawyers. The other contains records that describe firms. The records for lawyers contain information on the experience, compensation, and work habits of individual lawyers. The firm records contain information on the employee composition of different firms. Some individual lawyers appear in more than one annual survey because some firms are surveyed in multiple years. However, ALM did not provide identifiers that allow us to link these records over time.

Table 2 contains data on three lawyer-level variables: hours billed, hourly billing rate, and total compensation. We define total compensation as the sum of three measures collected by ALM; salary, bonus, and benefits. We express all monetary variables in 2011 dollars using the CPI-U as our inflation measure. This facilitates comparisons with the *After the JD* (AJD) data in Tables 3 and 4.

ALM collects data on five categories of lawyers: equity partners, non-equity partners, associates, counsel (of counsel) attorneys, and staff attorneys. Equity partners have ownership rights and control. Non-equity partners are lawyers that the firm presents to the public as partners even though these lawyers do not share the same capital contribution requirements, voting rights, or profit shares that full equity partners enjoy. Associates are under consideration for partner status. Counsels are not explicitly under consideration for a partnership, but they do tend to bill more than 800 hours per year. Staff attorneys are explicitly not being considered for a partnership. Our reading indicates that these attorneys are least likely to work full-time and enjoy the least employment security.

The firm level data reports totals for each type of lawyer employed in each firm. We sum these totals to create our firm size variable. Our measure of firm size is the sum of Full Time Equivalent lawyers in each firm.

We adopted several rules for cleaning the data. We corrected several obvious coding errors in the year barred variable.[37] When calculating compensation variables, we treat total compensation as missing if a lawyer (a) reports a salary below $10,000 or above $5,000,000 (b) reports a bonus greater than $5,000,000 (c) reports a benefit amounts less than $0 or greater than $100,000, or (d) reports less than 0 or more than 3,000 billable hours. We also code billing rates greater than $1,200 per hour as missing.

We calculate the experience of each attorney as the difference between the survey year and the year the attorney passed the bar. The ALM does ask lawyers to report their gender. We make little use of this variable since 50.70% of the lawyers in our data did not respond to this item.

After the JD. Tables 3 and 4 use data from the *After the JD* survey (AJD), conducted by the American Bar Association in three waves from 2002 to 2012. This longitudinal survey followed a stratified random sample of lawyers who were first admitted to the bar around 2000. The first stage of the two-stage sampling process divided the country into 18 strata based on the number of new lawyers in each area. In the second stage, researchers chose one primary sampling unit from each strata. These sampling units are local markets for legal services. The largest are the four "major" legal markets: Chicago, New York, Los Angeles, and Washington DC. These markets contain more than 2,000 new lawyers. Small states make up some of the smaller sampling units.

Within each primary sampling unit, researchers drew a random sample of individuals. They also drew an oversample of 1,465 new lawyers from minority groups. The final sample included 9,192 new lawyers. In wave 1, conducted from May 2002 to March 2003, 3,905 individuals from the national survey and 633 from the minority oversample responded, for a total of 4,538 respondents. Both are included in our sample. In wave 2, conducted from May 2007 through early 2008, researchers again reached out to the entire sample of lawyers, including those who had not responded in wave 1. In total, 4,160 respondents completed surveys in this wave. We do not use this wave in our analysis. In wave 3, conducted from May 2012 to December 2012, the ABA team surveyed those lawyers who had responded to either wave 1 or wave 2 or both. The wave 3 response rate was 53%, which created a sample of for 2,862 total respondents. 425 of these respondents were from the minority oversample.

We restrict the samples used in Tables 3 and 4 to respondents who

(1) respond to both the wave 1 and wave 3 surveys
(2) passed the bar in or after 1998

We divide our sample of lawyers into five categories based on the position that they reported in wave 3. Our categories include four positions in private law firms that are not run by solo practitioners. These positions are Partner, Non-Equity Partner, Associate, and Of Counsel (Counsels). We group all other lawyers in an Other category. This category contains persons who no longer work in a private law firm, solo practitioners, and a small number of contract or staff attorneys.

[37]Full list of changes: 205 became 2005, 208 became 2008, 1190 became 1990, 1194 and 1794 became 1994. Additionally, any year less than one hundred became that year plus 1900. There were no entries between 0 and 15, so it was appropriate to add 1900 in all cases rather than 2000

The ALM survey gave descriptions of different law firm positions in the survey instrument. The AJD does not provide definitions of the four positions we highlight. We argue in section 7 that most non-equity partners, especially those who have roughly 10 years of experience, are still trying to earn promotion to full equity partner and in some cases already function as partners in their interactions with clients. The AJD data provide support for this claim. All respondents to wave 3, not just the ones we select for our samples in Tables 3 and 4, provide a retrospective employment history. In wave 3, 1,472 lawyers report that they began their careers as associates in private law firms.[38] Only 92 of these lawyers report making partner in their original firm, and 18 of these lawyers report that they made partner after working as a non-equity partner in their original firm. Another 77 were promoted to non-equity partner in their original firm but had not made partner in the firm by wave 3. Of these, 22 left the firm before wave 3 and 55 remained.

The AJD data also support our contention that the transition from associate to counsel is not only less common but also often signals that the attorney in question is moving off the partnership track. In wave 3, only 40 the 1,472 lawyers who report beginning their careers as associates report employment as counsels in private law firms, and only 26 worked in counsel positions in their initial firms. Among these 26, only three moved back to the partnership track in their initial firms. One made partner and remained in the firm at wave 3. One made non-equity partner and later left. One made non-equity partner and stayed. Of the remaining 23, 11 left their initial firms and 12 remain in counsel positions at their initial firms in wave 3.

We define our key variables as follows: In wave 1, we use responses to the question "How many hours did you actually work last week, even if it was atypical?" to calculate average hours, and we use responses to the question "What is your total annual salary (before taxes) including estimated bonus, if applicable, at your current job?" to calculate the average salary. In wave 3, we use reports concerning the number of hours respondents are "Working at the office or firm (including being at court, clients' office, etc.) on weekdays," "Working from home on weekdays" "Working on the weekend," and "Attending networking functions" to calculate average hours. Our wave 3 compensation variable is the sum of reported values for "Salary," "Bonus," "Profit sharing/equity distribution," "Stock Options (present value)," and "Other."

We treat reported salaries of less than $10,000 as missing data. Wave 3 asks about compensation for calendar year 2011. We express all compensation measures from both waves in 2011 dollars using the CPI-U.

The AJD does not provide weights that adjust for differential attrition between waves 1 and 3. We did create versions of Tables 3 and 4 using the wave 1 sampling weights and found patterns that are quite similar to those in our unweighted analyses.

[38]The data for Tables 3 and 4 come from lawyers who responded to both the wave 1 and wave 3 surveys. This accounts for the significantly smaller sample sizes.