ALLOCATING EFFORT AND TALENT IN PROFESSIONAL LABOR MARKETS

GADI BARLEVY
FEDERAL RESERVE BANK OF CHICAGO
AND
DEREK NEAL
UNIVERSITY OF CHICAGO & NBER

PRELIMINARY AND INCOMPLETE -
PLEASE DO NOT CITE OR QUOTE WITHOUT PERMISSION

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 work loads while facing significant risks of dismissal. We argue that the productivity of skilled partners in professional service firms, e.g. law, consulting, investment banking, public accounting, etc, 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 work loads and up-or-out rules serve this purpose. Market participants learn more about new workers who perform more tasks, and 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. Over time, technological developments and evolving roles for specialists in large professional service firms may have shaped work hours and the degree of adherence to strict up or out rules in specific labor markets. We discuss how our model is able to rationalize these developments, and we also present evidence on life-cycle patterns of hours and earnings among lawyers that support key predictions of the model.

We thank Chris Taber, John Kennan, and Canice Prendergast for useful discussions. We thank seminar participants at UW-Madison, UCL, Edinburgh University, and the 2014 North American Summer Meetings of the Econometric Society for valuable feedback.

Date: November 23, 2015.

Introduction

Many professional service firms employ two personnel practices that are uncommon in other labor markets. First, these firms assign extremely heavy work loads to young professionals. Second, these firms often employ up-or-out promotion policies. These policies dictate that newly hired professionals expect to either progress to a position like equity partner within a relatively fixed number of years or leave the firm. They know that if the existing partners decide not to promote them to partner, they will also not give them the option to remain in a non-partner role.

In this paper, we develop a model that explains why many professional service firms require their new associates to work long hours while competing in up-or-out promotion contests. Our work fills a hole in the existing literature on professional labor markets. The literature on why young professionals work long hours does not overlap with the somewhat larger literature that seeks to understand why many of the same young professionals face up-or-out promotion policies. Rat race models and the literature on career concerns and reputation 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 this policy can be used to solve commitment problems or to remove workers who are ill-suited to professional work, but these models of commitment and screening often ignore worker effort and make no clear predictions about the efficiency of effort levels among new professionals.

While heavy work loads and up-or out promotion rules may serve many purposes in professional labor markets, our results suggest that they are not separate phenomena. Both heavy work loads 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. Market participants learn more about new workers who perform more tasks, and when firms replace experienced associates with new workers, they gain the opportunity to identify talented professionals who will have long careers as partners.

Both practices are costly. Work loads beyond statically optimal levels reduce the current surplus generated by new associates, and replacing competent, experienced associates with new associates lowers current output. However, in some professional labor markets, gains from increasing the number of talented professionals who occupy partner positions exceed both costs, and here, we expect to see both practices used in tandem.

In the next section, we review the literature on personnel practices in professional labor markets more carefully. Then, we present our model of work loads and job assignment. We show that both heavy work loads for young professionals and up-or-out promotion rules facilitate the identification of talented partners. In the penultimate section, we assess several predictions of our model using data from professional labor markets. 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

A significant literature documents the fact that new associates in elite professional service firms, e.g. leading firms in law, consulting, investment banking and public accounting, often work much longer hours than most white collar workers who have similar levels of education. Landers et al. [1996] offer a possible explanation for this pattern. They analyze data on law firms but suggest that their insights may apply to other professional labor markets as well.

Law firms and other professional services firms are often organized as partnerships. Building on the work of Akerlof [1976], Landers et al. [1996] treat law firms as teams and assume that, while team output is observed, individual output is not observed. In their model, teams share output 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. Thus, partners desire some screening mechanism that allows associates to reveal their type. 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. These contracts also describe auction mechanisms that dictate how existing partners in various firms will sell their ownership stakes to their associates in the future. 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. As in the canonical rat race model, Akerlof [1976], hours distortions are the equilibrium mechanism that sorts heterogeneous workers to heterogeneous teams. This model of hours requirements for new associates in professional labor markets does not 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 benefit from having strong reputations but 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 work more than efficient levels 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.[1]

In Landers et al. [1996], all associates are promoted, and no associates change firms or leave law as a profession. The model in Holmstrom [1999] contains only one job and therefore does not address promotion or retention decisions in any way.

[1]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.

While the literature on long hours among young professionals does not address promotion or retention decisions directly, the literature on up-or-out promotion rules has little to say about the long hours that young professionals work while participating in up-or-out promotion contests. The up-or-out literature contains several variations on two themes, but neither literature addresses why young professionals in up-or-out firms often work much more than other workers with similar levels of education.

Many papers characterize up-or-out rules as commitment devices that solve a double moral hazard problem between workers and firms. In these papers, 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. These rules force firms to dismiss all workers they do not promote and therefore punish firms if they make unfavorable reports about workers who produce positive output signals. The existence of this punishment allows firms to credibly promise to reward hidden actions, and this credibility allows firms to elicit more efficient actions from workers.

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. Prendergast [1993] argues that up-or-out rules are not always needed to solve the double moral hazard problem that Kahn and Huberman [1988] identify. If firms have positions for skilled workers that sufficiently leverage their skills, they incur costs when they fail to make deserved promotions, and these costs may make contingent promises concerning raises and promotions credible. Thus, firms can induce workers to invest in firm-specific skills without employing up-or-out rules as long as the firm benefits from promoting all workers who do invest.

Waldman [1990] extends the logic of Kahn and Huberman [1988] to an environment where firms have private information about worker ability as opposed to skill investments. He shows that private information about worker talent creates the same moral hazard problems that Kahn and Huberman [1988] describe even when all human capital is completely general. When firms have private information about how productive their workers would be in other firms, they are tempted to deny promotions to deserving workers in order to maintain their information rents. In this scenario, firms may use up-or-out rules as commitment devices, and by committing to more efficient promotion decisions, they induce young workers to invest more efficiently in skills.

Ghosh and Waldman [2010] extend Waldman [1990] to an environment where new professionals take hidden actions that influence output signals and a portion of worker productivity is firm-specific. They conclude that up-or-out is more likely in professional labor markets where the promotion of workers to senior positions has relatively small effects on their productivity and most human capital is not firm-specific. In these settings, the ex post surplus generated by efficient promotions is relatively low. Therefore, firms demand a mechanism that allows them to credibly commit to contingent promises concerning raises and promotions.[2]

[2]Ghosh and Waldman [2010] also model worker effort. However, in contrast to our model, worker actions are hidden in their model. As in Holmstrom [1999], workers may expend effort to

A separate literature on up-or-out promotion rules links up-or-out rules to optimal screening procedures. OFlaherty and Siow [1992] develop a model of professional partnerships where partners work with one associate and receive signals about the suitability of the associate for promotion. They argue that partnerships grow by identifying people who are talented enough to be partners and show that the optimal screening rule in their environment involves two cutoffs. When the posterior belief about an associate crosses the upper cutoff, the candidate becomes a partner and takes on a new associate. When the posterior falls below the lower cutoff, the existing associate is dismissed and replaced by a new associate. Since beliefs about all associates eventually cross one of these thresholds, each associate either goes up or out.

Demougin and Siow [1994] link up-or-out rules to screening 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.

OFlaherty and Siow [1992] and Demougin and Siow [1994] describe an up-or-out equilibrium where new professionals are no more productive than the experienced professionals they replace, but these new professionals have more option value than their more experienced counterparts. Below, we derive results that link up-or-out rules with this same option value logic. Yet, our results differ in several ways. First, we model worker effort and introduce a signaling technology such that the market learns more about new professionals when these 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 in professional labor markets 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 attractive while changes that raise the relative productivity of partners make up-or-out rules more productive.

Third, our model provides new insights concerning the interpretation of outcomes in up-or-out firms. 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

influence signals that determine their reputation, but this is true in firms that employ standard promotion practices as well as those that adopt up-or-out rules. Further, effort levels among new professionals may be above or below efficient levels in both up-or-out firms and firms that follow standard promotion practices.

partners. If these young professionals learn they are not going to become partners, they are no longer willing to pay this cost.

Our work does not represent a direct challenge to the existing theoretical literature on up-or-out. We assume that all market participants in a given professional labor market learn symmetrically about all other market participants, but we do not address the costs of verifying information for courts. Thus, nothing in our work eliminates the possibility that up-or-out rules do help firms solve important commitment problems.[3] In addition, our results support a key idea in both OFlaherty and Siow [1992] and Demougin and Siow [1994], since we too conclude that up-or-out rules should be interpreted as optimal screening procedures.

Yet, our model produces several new insights about professional labor markets. We show how both heavy work loads and up-or-out promotion rules facilitate searches for talent, and our framework produces new comparative static results concerning the joint determinants of work loads for young professionals and the existence of up-or-out promotion rules. Finally, we produce predictions about life-cycle patterns of changes in hours worked for professionals who follow different career trajectories, and these predictions appear to match data from several sources.

2. Model Setup

Our model describes production, learning, and job assignment in professional labor markets. Individuals may work in the professional sector or in 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. The return to ability is greatest in the partner position. We use the term partner because professional service firms use this title for their leaders. However, we do not model the organizational structure of these firms. In our model, the term partner simply refers to a position where the most skilled professionals are most productive.[4]

We seek to understand how markets learn about workers and make job assignments. In our framework, learning is public, there are no hidden actions, and there are no firm-specific skills. Given this setting, it is convenient to describe our results as solutions to a learning and assignment problem that faces a benevolent social planner. After we present our results, we describe a market for professional services. The competitive equilibrium in this market produces the allocations that solve our planner's problem.

Given our assumptions, the work effort of new associates in the professional sector produces output, and the relationship between effort and output provides information about the ability of new associates. If their work reveals that they

[3]Levin and Tadelis [2005] suggest the up-or-out helps partnerships solve a commitment problem with their customers. Professional service firms promise to supply talented professional who perform quality work, but clients of professional service firms may find it difficult to judge the talent of different professionals ext ante. Up-or-out rules make promises concerning the quality of professional services more credible because clients of a given up-or-out firm know that the other partners in the firm have agreed to share revenue with the partner directing the work on their case.

[4]In public accounting, investment banking, and consulting firms, the path to partner involves more than two job titles. Here, we use the two titles that are common in law firms that have the simplest structure. We do so for convenience. A model with many potential up-or-out decisions points on the path to partner would be much more cumbersome, but the basic insights from our model would remain.

are talented, the planner assigns them to be partners, and their value increases. If their work reveals that they are not talented, the planner assigns them to work outside the professional sector. If their work fails to reveal their true talent level, the planner never assigns them to the partner position, but he may retain them to work as senior associates. If the planner instead reassigns all these associates to jobs outside the professional sector, we say that the planner follows an up-or-out assignment rule. In these environments, the planner promotes associates if their first-period output reveals that they have high ability, and the planner assigns all other associates to positions outside the professional sector.

In our model, new associates are options, and their option value is determined by the gains associated with identifying talented professionals and promoting them to partner. We show that up-or-out promotion rules are optimal when the productivity of talented partners is great enough, relative to the expected productivity of senior associates who possess average ability. Further, in these same markets, heavy work loads for new associates are also optimal because they reveal more information about associate talent levels and therefore permit the identification and promotion of more new partners.

We begin by describing preferences, production, and the learning technology. Next, we derive optimal policies concerning effort levels and job assignments. Finally, we endogenize the size of the professional sector and show that simple decentralized mechanism implements the planners choices for effort levels, job assignments, and sector size.

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.

Workers are risk neutral with the following utility function

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

where x 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, $\lim_{n\to\bar{n}} c'(n) = \infty$, $c''(n) > 0 \quad \forall n \in [0, \bar{n}]$. Below, we describe a planner's problem where the planner assigns workers to jobs and work loads, n. These work loads are the number of tasks that the planner assigns to each worker. 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 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, but no one has private information about their own ability or the ability of others.

There are two sectors in the economy. In the outside sector, output, y^o, is a deterministic, linear function of worker effort, and the mapping between effort and output does not vary with worker experience or ability. The production function in the outside sector is

$$y^o = w^o n$$

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. In contrast to the outside sector, output is stochastic in the professional sector. Nature draws i.i.d. production shocks, ϵ, that are mean zero for all professional workers in each period. The production function for new associates is

$$y_0^a = (1+\theta)n + \epsilon \tag{2.1}$$

The production function for experienced associates is

$$y_1^a = z^a(1+\theta)n + \epsilon \tag{2.2}$$

Here, the parameter $z^a > 1$ captures the idea that associates who have experience are able to perform more productive tasks.

Finally, the production function for partners is

$$y_1^p = \begin{cases} z^p(1+\theta)n + \epsilon & \text{if } \theta = x \\ -\infty & \text{if } \theta = 0 \end{cases} \tag{2.3}$$

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 $y_0^p = -\infty$ for all workers. This assumption captures the idea that, regardless of their ability, workers with no experience would also make costly mistakes if they were to act as partners.

Our planner must allocate workers between the professional labor market and all other employments. For now, we cap employment in the professional sector at $q < 1$ to capture the idea that only a fraction of highly-educated workers begin their careers in the professional sector. 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 are interested in assignment decisions. These decisions involve interesting trade-offs if the following productivity relationships hold

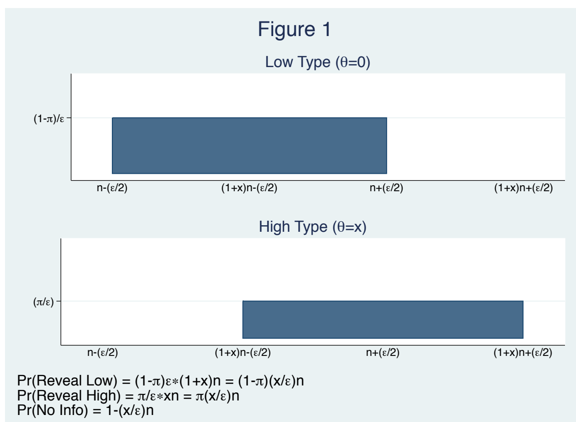
$$z^a < w^o < z^a(1+\pi x) \tag{2.4}$$

The first inequality in equation 2.4 implies that an experienced associate who has low ability is more productive in the outside sector than the professional sector. To understand the second inequality, recall that π is the probability that a given worker has high ability, $\theta = x$, and z^a captures how productivity grows with experience among those who remain in the associate position. Thus, the second inequality implies that the expected productivity of an experienced associate with unknown ability is greater than both the expected productivity of a new associate and the productivity of labor in the outside sector.

No one observes worker ability directly, and no one possesses private information about the ability of any worker. However, everyone observes the work load, n, and resulting output for each worker, y_s^j.

2.2. **Learning.** We follow Pries (2004) and assume that the production shocks are uniformly distributed, $\epsilon \sim U[\frac{-\varepsilon}{2}, \frac{\varepsilon}{2}]$. This implies that learning in our model has an "all or nothing" feature. The output signal for a given new associate either reveals the associate's ability perfectly, or it reveals nothing.

Figure 1 illustrates why learning has this all or nothing feature. The two panels in the figure present two joint densities. Both densities describe output realizations for an associate who takes on a given work load, 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.



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

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

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. It follows immediately that $\pi\frac{x}{\varepsilon}n$ is the probability that the output signal reveals that an associate has high ability, and $(1-\pi)\frac{x}{\varepsilon}n$ is the probability that the signal reveals that an associate has low ability. We assume that $\frac{x}{\varepsilon}\bar{n} < 1$ 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 work loads do create more information in this model. Since $x > 0$, the lower bound of output for high-ability workers grows faster with n than the upper bound of output for low-ability workers. So, as n increases, the region of overlap in Figure 1 shrinks. New associate effort, n, not only produces output but also reveals information new associates.

Given our assumptions on the production technologies, 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, but optimal second period assignments for associates of uncertain ability are more subtle. Our analysis below highlights how z^p, z^a, x, π, and ε interact to determine both optimal work loads for new associates and whether or not experienced associates of uncertain ability face dismissal.

3. The Planner's Problem

Here, we describe the planner's problem for our economy. The planner seeks to maximize the present discounted value of the the sum of present and future differences between per period output and effort costs by assigning workers to jobs and work loads.

In each period, the planner's problem involves ten choices. Table 1 demonstrates that there are five different types of workers in this economy. The planner must choose a job assignment and effort level for each type. We proceed by showing that the optimal job assignment for three of these types is immediate. We then argue that the optimal effort levels for four of these five types are solutions to straightforward static optimization problems. Thus, 7 of the planner's 10 choices are immediate given our assumptions. We devote our analysis below to the two assignment decisions and one effort choice that remain.

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.

The rows of Table 1 describe three different types of workers with respect to their previous work experience. Recall that workers live two periods, and there are two sectors. Thus, workers may have no experience in either sector, one period of experience in the outside sector and no experience in the professional sector, or no experience in the outside sector and one period of experience in the professional sector. The columns of Table 1 describe three different information states that may apply to workers. The market may know that a worker is low ability, $\theta = 0$. The market may know the worker is high ability, $\theta = x$, or the market may be uncertain about the worker's ability and believe that there is a probability π that the worker is high ability.

The intersections of these three experience types and three information sets yield nine cells in Table 1. We begin by explaining why the first two columns of the first two rows are marked, n.a., for not applicable. These rows describe workers who have no professional experience yet are known to be either high or low ability. Since all workers are born with uncertain ability, and all learning takes place in the professional sector, no one can know the true ability of any worker who has no professional experience. Thus, these four cells describe types that never exist. The five cells in the last column and last row of Table 1 describe types that may exist. The planner must assign these types to jobs and work loads.

Three of the planner's five job assignment decisions are trivial. First, new workers in the top right corner of Table 1 are of uncertain ability. So, they never work as partners. Given our linear production technologies, the planner either assigns all new workers to the outside sector, or he assigns new workers to associate positions in the professional sector until the constraint on professional employment, q, binds and then assigns the remaining new workers to the outside sector. We assume that the later case holds. If not, the planner never assigns new workers to the professional sector, and all workers spend their entire career in the outside sector.

Next, consider workers in the bottom left hand corner. These workers have one period of professional experience as an associate, and their output signals have revealed that they are low ability. The planner clearly assigns these workers to the outside sector. Because they have only one period of life remaining, option value considerations cannot affect their assignment, and by assumption, they are more productive in the outside sector, since $z^a < w^0$.

Finally, turn to the second row and last column. These workers have one period of experience in the outside sector. It is straightforward to see that the planner keeps these workers in the outside sector. To begin, the planner never assigns these workers to be partners in the professional sector because their abilities are uncertain. In addition, the planner never assigns these workers to associate positions. Each new worker has the same expected associate productivity as an experienced outside worker plus the prospect of being promoted to partner in the next period, and since employment in the professional sector is capped at $q < 1$, the planner could fill the whole sector with new associates.

This leaves two job assignment decisions for the planner. The cells in the second and third columns of the bottom row contain professional workers who have one period of experience. Recall that once the planner decides how many experienced professionals to retain in the professional sector, the planner fills the remaining slots in the professional sector by assigning new workers to the associate position. Thus, the mix of new associates and new outside workers in the top right cell is pinned down by the assignments of experienced professionals.

In the following section, we characterize optimal assignment rules for experienced professionals. It is relatively straightforward to show that the planner promotes high ability types, $\theta = x$, to partner. However, the optimal assignment rule for experienced professionals of uncertain ability is more nuanced.

Next, consider optimal work loads for the five types described in the bottom row and last column of Table 1. Four of these five types are experienced workers. These workers have only one period of life remaining, so the planner assigns them work loads that equate expected marginal products of effort with the marginal costs of effort.

Finally, the planner must assign work loads to the new workers in the top right hand corner of Table 1. Among those who begin in the outside sector, output levels produce no information about worker ability, and we have shown above that the planner will not assign these workers to the professional sector in next period. Thus, optimal effort for these workers also corresponds to the solution of a simple static maximization problem.

In contrast, the optimal effort choice for new workers who start their careers as associates in the professional sector is more interesting. Among these workers, heavier work loads generate more output and more information about worker ability. Because this information guides assignment decisions in the next period, the optimal work load for new associates must reflect the fact that effort today affects assignments and work loads in the future.

To review, our planner's problem involves five job assignment decisions and five effort choices. In seven of ten cases, the planner's optimal policies are immediate. Further, we quickly establish below that the planner always assigns experienced professionals with known high ability to the partner position. Thus, our planner must confront two key questions: Should experienced associates whose abilities remain uncertain continue working as associates in the professional sector, and what is the optimal work load for new associates? In the following section, we analyze these questions and demonstrate how the answers to these two questions are connected.

3.1. **Recursive Formulation.** In this section, we describe the planner's optimal policies as solutions to a specific Bellman equation. Before describing this equation, we introduce some additional notation.

- v^o is the per period surplus created by an outside sector worker.

Our assumptions imply that all workers in the outside sector produce the same amount and incur the same effort costs regardless of their past work experiences or talent level.

For professional workers, v_s^j describes the per period surplus created by a professional worker with $s = \{0, 1\}$ periods of professional experience in position $j = \{a, p\}$.

- $v_0^a(n)$ - the expected surplus created by a new associate who takes on a work load of n.
- v_1^a - the expected surplus created by an experienced professional of uncertain ability who works as an associate
- v_1^p - the expected surplus created by an experienced professional of high ability who works as a partner

We omit explicit notation for effort, n, in v^o,v_1^a, and v_1^p, because optimal effort levels for these workers are solutions to simple, static maximization problems. We include n in $v_0^a(n)$ because the effort of new associates both produces output and creates information that influences future payoffs. Our notation does not specify beliefs about the abilities of experienced associates or partners because, given our production function assumptions, the planner only assigns workers with known high-ability to the partner position, and the planner only retains experienced workers of uncertain ability as associates.

Now, consider the stock variables for our problem. Each period, there is a mass one of new workers. There are experienced workers who spent their first period of life in the outside sector, and there are three types of experienced workers who spent their first period of life in the professional sector: those with known low ability, those with known high ability, and those with uncertain ability.

In our formulation, we explicitly track only two of these five stock variables.

- ρ^u - the mass of experienced professionals who have uncertain ability, $Pr(\theta = x) = \pi$
- ρ^x - the mass of experienced professionals who have known high ability, $\theta = x$

Recall that the planner retains all experienced outside workers in the outside sector and also assigns all professionals with known low ability to the outside sector. This implies that the mass q of workers in the professional sector is divided among new associates and two types of experienced professionals. The planner's assignment decisions concerning the stocks of experienced professionals with uncertain ability or known high ability pin down the mass of new associates in the professional sector, and this pins down allocations of new workers are among the outside and professional sectors.

The key control variables for our planner involve job assignments for experienced associates who are not known to have low ability and the work load assignment for new associates.

- α^u - the fraction of experienced professionals with uncertain ability, ρ^u, that the planner retains in the professional sector, $\alpha^u \in [0, 1]$

- α^x - the fraction of experienced professionals with known high ability, ρ^x, that the planner retains in the professional sector, $\alpha^x \in [0, 1]$
- n - the work load or effort level for new associates

The per period surplus flow in this model is

$$s(n, \alpha^u, \alpha^x) = (2 - q)v^o + (q - \alpha^u \rho^u - \alpha^x \rho^x)v_0^a(n) + \alpha^u \rho^u v_1^a + \alpha^x \rho^x v_1^p \tag{3.1}$$

The control variables (α^u, α^x) pin down the entire allocation of workers to positions. Given an allocation of workers to positions, work loads determine expected output. Four of these work load decisions are trivial, but the work load, n, for new associates not only influences expected output but, given the current stock of new associates, also determines the key stock variables, ρ^u and ρ^x, next period. Thus, the planner's policies concerning α^u, α^x, and n drive the evolution of stocks and output flows over time.

The choice variable α^x does not explicitly involve the choice to assign high-ability workers to the partner position. However, we note above that the condition, $z^p > z^a > 0$, implies that whenever the planner retains an experienced professional with known high ability in the professional sector, the planner also assigns this worker to the partner position. Likewise, the variable α^u does not explicitly involve the choice to assign the uncertain-ability professional to the associate position, but our assumptions about productivity in the partner position ensure that professionals of uncertain ability are never promoted to partner.

The planner's objective is to choose efforts levels and job assignments that maximize the discounted present 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

$$\begin{aligned} V(\rho^u, \rho^x) \quad = \quad & \max_{\alpha^u, \alpha^x, n} \quad s(n, \alpha^u, \alpha^x) + \\ & \beta V\left([q - \alpha^u \rho^u - \alpha^x \rho^x] \left[1 - \frac{x}{\varepsilon} n\right], \, [q - \alpha^u \rho^u - \alpha^x \rho^x] \, \pi \frac{x}{\varepsilon} n \right) \end{aligned} \tag{3.2}$$

The most straightforward way to understand this equation is to consider the special case $V(0, 0)$. In this case, no experienced professionals with high or uncertain ability exist, so the planner assigns all new workers to the associate position in the professional sector and all older workers to the outside sector. Surplus this period, $s(n, \alpha^u, \alpha^x)$, equals $(2 - q)v^o + qv_0^a(n)$. At the end of the next period, the outside workers expire, and the stock of experienced associates with uncertain ability is $q(1 - \frac{x}{\varepsilon}n)$ while the stock of experienced associates with known high ability is $q\pi\frac{x}{\varepsilon}n$.

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, $\rho^u > 0, \rho^x > 0$, 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 $\frac{x}{\varepsilon}n$.

There are no productivity spillovers among workers who occupy different positions in this model, and the output generated in one position is not a function of total employment in the position. Thus, we quickly establish below that $V(\rho^u, \rho^x)$ is linear. Given this result, we easily establish that the planner always chooses $\hat{\alpha}^x = 1$.

3.2. **Promotion to Partner.** In Appendix A, we demonstrate that $V(\rho^u, \rho^x)$ defined in equation 3.2 is a contraction mapping. Given this result, we then establish the following claim:

Claim 1. $V(\rho^u, \rho^x) = K_1 + K_2\rho^u + K_3\rho^x$ for some constants K_1, K_2, and K_3.

This claim establishes that $V(\rho^u, \rho^x)$ is linear in the two stocks of experienced professionals, which implies 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 the stock. This feature of our model implies our second claim.

Claim 2. $\hat{\alpha}^x = 1$

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, $\pi\frac{x}{\varepsilon}\hat{n} < 1$, 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. Further, as we note above, the planner always promotes them to partner when retaining them, since $z^p > z^a$.

We have now determined optimal rules for eight of the ten choices the planner makes each period. There are two choices that remain: the work loads for new associates, n, and the sector assignment for experienced associates of uncertain ability, α^u. Recall that this assignment decision is really a retention decision. The planner never allows professionals of uncertain ability to work as partner, so the decision to retain an experienced professional of uncertain ability in the professional sector is equivalent to the decision to retain her as an associate.

4. The Link Between Work Loads and Up-or-Out

We have reduced our planner's problem to two choices. The planner must choose the work load for new associates, and the planner must choose whether or not experienced associates of uncertain ability remain in the professional sector. Here, we demonstrate how these choices are related.

Substitute the expression for $V(\rho^u, \rho^x)$ in Claim 1 into 3.2. Then, take the derivative with respect to n to get the first order condition that defines the optimal work load for new associates, $\hat{n}$:

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

This equation highlights an important property of optimal work loads for new associates. The marginal cost of new associate effort must equal the sum of two marginal returns. The first, $(1+\pi\theta)$, is the expected marginal product of new associate effort. The second, $\beta\frac{x}{\varepsilon}(\pi K_3 - K_2)$, is the marginal information return from worker effort. Note that K_2 and K_3 represent incremental values to the planner of replacing a new associate with an experienced professional who possesses either uncertain ability or known high ability, respectively. Our learning technology implies that when the planner increases a new associate's work load by Δn, the probability that the associate's output signal reveals her ability increases by $(\Delta n\frac{x}{\varepsilon})$. 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, $\frac{x}{\varepsilon}(\pi K_3 - K_2)$ 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, and we also prove the following in Appendix A:

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

New associates work long hours that reduce the current surplus generated by their positions because these long hours 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. Given 3.2, the first order condition that defines $\hat{\alpha}^u$ is

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

In Appendix A, we solve for K_2 and K_3 and use these solutions to express the first order condition for α^u in a more informative way.

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

The planner imposes an up-or-out rule, i.e. he promotes experienced associates with known high ability to partner and assigns all other experienced associates to the outside sector in situations where the expected future return from adding a new associate this period is greater than the expected returns to experience among associates of uncertain ability.

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. This cost is positive for two reasons. Experienced associates are more productive than new associates, and new associates work more than the statically optimal level. The right-hand side gives the expected future returns that these replacements create. Experienced associates with uncertain ability are never going to be partners, but new associates may become partners next period. The probability of this event is $\beta\pi(\frac{x}{e})\hat{n}$, and the additional surplus generated by known high-ability professionals who work as partners is $v_1^p - v_1^a$.

Thus, the planner's decision concerning $\hat{\alpha}^u$ reflects a trade off between the return to worker experience in the associate position and the value of identifying high-ability workers and promoting them to the partner position. As in our discussion of heavy work loads among new associates, up-or-out rules reduce current surplus but increase expected surplus in the future by increasing the number of high-ability professionals who become partners.

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 $\hat{\alpha}^u$. Appendix A proves the following result concerning optimal work loads:

Proposition 3. *The optimal work load for new associates, $\hat{n}$, is increasing in z^p and non-increasing 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 above that many professional service firms employ both heavy work loads 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 work loads for new associates. Our second comparative static result confirms this:

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

This result follows because the cost of letting an experienced professional of uncertain ability go is increasing in z^a, and the returns from finding talented professionals and promoting them to partner is increasing in z^p.

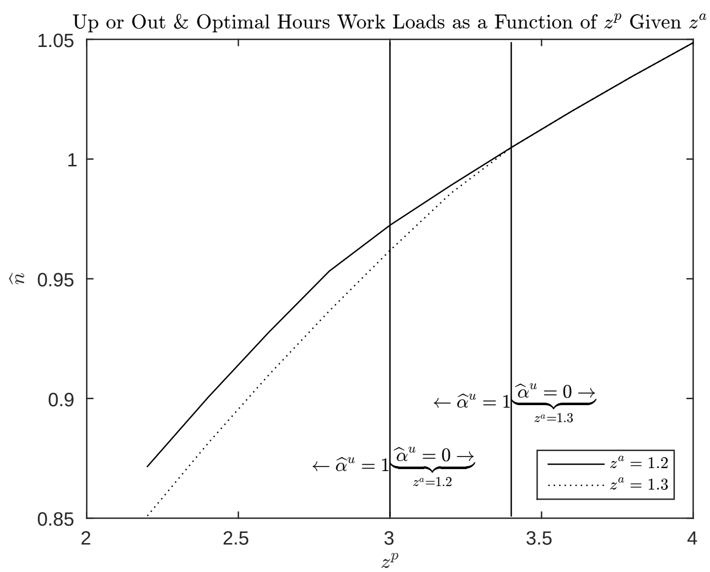
Figure 2 traces out $\hat{n}$ and $\hat{\alpha}^u$ as functions of z^p given two different values of z^a and holding the cost function and other model parameters fixed. This figure highlights the sense in which up-or-out promotions rules and heavy work loads for new associates go together. In both scenarios, $\hat{n}$ increases with z^p. Further, given values of (z^a, z^p) such that optimal policies do not involve up-or-out, $\hat{\alpha}^u = 1$, one can always create an up-or-out equilibrium, $\hat{\alpha}^u = 0$, by increasing z^p. Likewise, holding z^p constant, $\hat{n}$ is weakly decreasing in z^a, and given values of (z^a, z^p) such that an up-or-out rule is optimal, $\hat{\alpha}^u = 0$, one can always create an $\hat{\alpha}^u = 1$ equilibrium by increasing z^a.

In environments where z^p is large enough relative to z^a, the planner chooses heavy work loads 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.

In elite professional service firms, those who achieve the rank of partner, convince clients to allow them to direct projects where millions and possibly billions of dollars hinge on the quality of their decisions. Although many market mechanisms help match the best young professionals with these firms, important differences in true talent remain among the cohorts of new associates who enter these firms. Our approach is built on the premise that these elite professional service firms learn about these talent differences by requiring new associates to perform work. Given this starting point, we argue that up-or-out as well as heavy work loads for new associates are mechanisms that firms employ to identify the talented professionals who can attract and maintain valuable clients.

Gorman [1999] points out that strict adherence to up-or-out rules became less common in law firms during the 1980s and 1990s. She argues that, as the market for legal services grew, private law firms began doing relatively more work that required special expertise and experience. Thus, experienced, skilled lawyers who were not well-suited to the role of partner became more valuable. If we phrase these claims in terms of the language of our model, Gorman [1999] is arguing the z^a increased over time and law firms responded by moving away from strict up-or-out rules. As the market for professional services has grown in recent decades, we conjecture that growth in the relative demand for specialists has produced similar movements away from strict up-or-out policies in other professions as well.[5] Nonetheless, many new associates in elite professional service firms still begin their careers expecting that, within the coming decade, they will either move up to partner or out to another employer.

[5]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. See **New York Times**, May 17, 1990.


Up or Out & Optimal Hours Work Loads as a Function of z^p Given z^a
$\widehat{n}$
1.05
1
0.95
0.9
0.85
2
2.5
3
3.5
4
z^p
$\leftarrow \widehat{\alpha}^u = 1 \underbrace{\widehat{\alpha}^u = 0 \rightarrow}_{z^a=1.3}$
$\leftarrow \widehat{\alpha}^u = 1 \underbrace{\widehat{\alpha}^u = 0 \rightarrow}_{z^a=1.2}$
$z^a = 1.2$
$z^a = 1.3$

4.1. **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 the incumbent worker. Thus far, we have assumed $z^p > z^a$, but if $z^p = z^a$, our model becomes a one position model. 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.

Although some may describe the more stringent screening rule, $\hat{\alpha}^u = 0$, as an up-or-out policy, 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 productivity of different positions in an organization.

Partners in professional service firms are responsible for the business development and client management activities that allow these firms to survive and grow. Thus, skilled professionals who not only do excellent professional work but are also able to develop and maintain productive relationships with clients create great value for professional service firms. On the other hand, many professional service firms do not create productive roles for skilled, experienced, professionals who do quality work but are not able to develop and maintain new clients.

Things are different in large manufacturing, retail, or traditional service firms. Young professionals enter these firms knowing that they may end up in one of many different positions that will leverage the talents and expertise that they discover and develop over time, and most of these positions are not the highest positions in these firms. These firms have many productive roles for skilled professionals who are not well-suited to these leadership roles, e.g. CEO, CFO, etc.

5. Endogenous Sector Size and Decentralization

Up to now, our planner has solved allocations problems subject to the constraint the professional employment cannot exceed a cap of q, and we have assumed production technologies such that this cap on professional employment always binds. Here, we show that our main results hold in a more general setting where the planner also chooses the optimal size of the professional sector. Further, we characterize an equilibrium of a decentralized economy that implements the solution of this more general planner's problem.

5.1. **Endogenous Sector Size.** We assume that maintaining a professional position is costly. Professional workers require support staff that facilitate their capacity to interact efficiently with clients. Further, we assume that the supply curve for effective support services is upward sloping because each potential support worker has a different outside option in another sector. Thus, the resource cost of supporting a given professional position is determined by the outside option of the staff who support that position.

Suppose the per-period cost of supporting the q-th position is given by $\kappa(q)$, and $\kappa(\cdot)$ is an increasing function where $\lim_{q\to 0} \kappa(q) = 0$ and $\lim_{q\to 1} \kappa(q) = \infty$. These restrictions ensure that some but not all of the workers in a cohort begin their careers in the professional sector. The planner faces the same problem as before, but now, he must also determine $\hat{q}$, the optimal number of professional workers.

Recall 3.1 and note that expected surplus in the current period depends on q, which is now a choice variable for the planner. The new planner's problem is

$$\begin{aligned} V\left(\rho^u, \rho^x\right) &= \max_{n,\alpha^u,\alpha^x,q} \quad s(n, \alpha^u, \alpha^x, q) - \int_0^q \kappa(y)dy + \\ &\quad \beta V\left(\left[q - \alpha^u\rho^u - \alpha^x\rho^x\right]\left[1 - \frac{x}{\varepsilon}n\right], \left[q - \alpha^u\rho^u - \alpha^x\rho^x\right]\pi\frac{x}{\varepsilon}n\right) \end{aligned}$$

It is straightforward to show that the three first-order conditions for n, α^u, α^x are the same as the first-order conditions in our original formulation. In particular, these conditions are independent of the size of the professional sector. As a result, $V(\rho^u, \rho^x)$ remains linear, and all of our results continue to hold. The planner's optimal sector size choice does not impact optimal work loads for new associates or optimal retention or promotion rules. Rather, the planner's choice, $\hat{q}$, is pinned down by the expected surplus that these optimal policies generate and the shape of $\kappa(q)$.

5.2. **Decentralization.** We next turn to the question of whether the planner's allocation can be achieved in a decentralized market economy. Observe that our model does not involve any information asymmetries: Workers have no private information about their abilities or their actions, and all output signals and all actions are public. Thus, it is many different market mechanisms could implement the solution to our planner's problem. Appendix B proves that one particular mechanism does. Here, we discuss how and why this mechanism would work.

The most straightforward way to decentralize our planner's problem is to 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, i.e. they choose whether to work as an associate or as a partner and then receive the output they create. 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 work loads and wages equal to expected marginal products for all possible combinations of worker types and positions.

In this decentralization, each professional must hire support services at a cost $\kappa(\tilde{q})$, where $\tilde{q}$ is the number of workers in the professional sector.[6] In this setting, a new worker who enters the outside sector receives lifetime utility $(1+\beta)v^o$, and a new worker who enters a professional sector of size $\tilde{q}$, works as an associate, and chooses workload n expects to enjoy the following expected lifetime utility:

$$v_0^a(n) - \kappa(\tilde{q}) + \beta\pi\frac{x}{e}n(v_1^p - \kappa(\tilde{q})) + \beta(1-\pi)\frac{x}{e}nv^o + \beta(1 - \frac{x}{e}n)\max[v^o, v_1^a - \kappa(\tilde{q})]$$

Here, we are imposing that professional workers expect the size of the professional sector to remain constant in both periods of their life. The first term of this expression reflects that fact that workers never choose to work as professional partners unless they know they are high ability because they would suffer an infinite penalty if they acted as partners and actually possessed low ability. Further, if an associate realizes that she has low ability, she always switches to the outside sector

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

in her second period, given our production assumptions. Finally, experienced associates who are still uncertain about their ability choose between the outside sector and continuing to work in the professional sector.

Recall that $(\hat{n}, \hat{q}, \hat{\alpha}^u)$ are the planner's solutions to our allocation problem, and also recall Proposition 3 that characterizes $\hat{\alpha}^u$. Then, consider the following three conditions:

$$(1+\beta)v^o = v_0^a(\hat{n}) - \kappa(\hat{q}) + \beta\pi\frac{x}{e}\hat{n}(v_1^p - \kappa(\hat{q})) + \beta(1-\pi)\frac{x}{e}\hat{n}v^o + \beta(1-\frac{x}{e}\hat{n})\max[v^o, v_1^a - \kappa(\hat{q})]$$

$$\hat{n} = \arg\min_n \quad v_0^a(n) - \kappa(\hat{q}) + \beta\pi\frac{x}{e}n(v_1^p - \kappa(\hat{q})) + \beta(1-\pi)\frac{x}{e}nv^o + \beta(1-\frac{x}{e}n)\max[v^o, v_1^a - \kappa(\hat{q})]$$

$$v_1^a - v_0^a(\hat{n}) < \beta\pi(\frac{x}{e})\hat{n}(v_1^p - v_1^a) \Longleftrightarrow v^o > v_1^a - \kappa(\hat{q})$$

The first states that new workers are indifferent between beginning their careers in the professional sector or the outside sector. The second states that new associates choose the planner's solution for first period effort as their optimal work load. The final condition states that experienced associates leave the professional sector and choose outside work if and only if the production environment is such that the planner would make the same assignment. Appendix B shows that all three of these conditions hold, which means that a competitive equilibrium exists that implements the planner's solution.

One aspect of our competitive equilibrium merits particular attention. Whenever the planner's solution dictates an up-or-out rule for professional workers, either z^a is so small or $\kappa(\hat{q})$ is so large that experienced associates of uncertain ability reject professional employment and choose the outside sector contract instead. Appendix B demonstrates the following:

Proposition 5. *Holding z^a fixed, $\hat{q}$ and $\kappa(\hat{q})$ are increasing functions of z^p.*

When partners become more productive, the professional sector grows, and the costs of maintaining professional positions grows. Our decentralization shows that when up-or-out is optimal, z^p, z^a, and the resulting $\kappa(\hat{q})$ are such that, all associates, who realize they are not going to become partners, willing choose the outside sector. In up-or-out regimes, the search for talented partners drives the costs of maintaining professional positions so high that each position is occupied by someone who either is a partner or could become one.

New associates willingly accept lower expected utility in their first period of employment because they are willing to pay to learn about their talent levels. Among experienced professionals, there is no next period, and therefore information gained this period has no value.

The fact that working as a new associate involves paying a up-front utility cost in exchange for the possibility of becoming a partner may shed light on survey evidence concerning the job satisfaction of young professionals. Young professionals often report low job satisfaction, and in particular, they report that they would be willing to accept lower earnings in exchange for less demanding work schedules. While advocates of rat race models cite these responses as evidence that young professionals take on work loads that are inefficient, our model offers another interpretation.

Assume that young professionals who respond to such surveys are reporting that, holding all else constant, they are willing to accept lower earnings in exchange for less demanding work loads. Further, assume that one of the things they hold constant when answering these questions is their future prospects for promotion. Given these assumptions, new associates in our model would express the same willingness to exchange current salary for reduced work loads. 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.

Although many young professionals report that, relative to their current terms of employment, they would be willing to exchange money for leisure, these reports are not direct evidence of market failure. Information is costly, and these reports may simply mean that workers would rather live in a world that allowed them to discover and reveal their abilities at no cost.

Scholars have argued that up-or-out policies are puzzling because, when a professional service firm decides not to promote a given associate, the firm often helps the associate acquire a job with one of its clients. The firm would not engage in this placement activity if it did not believe the associate was a competent professional. So, previous work has sought to understand why these firms do not make retention offers and employ these skilled associates in a non-partner role.

In our decentralization, associates who reach the end of the first period and have uncertain ability have the opportunity to remain in the professional sector at a higher wage rate than the rate they enjoyed as associates, but they always choose the outside sector contract instead. Professional sector workers must pay the market price for the resources that support their positions, and in up-or-out environments, only those who are partners or have a chance to become partners are willing to pay this cost.[7]

Our model suggests that once a young professional learns that she is not "partner material," her options in the corporate sector dominate any retention offer that her firm would make. This makes sense since large corporations offer numerous positions that allow skilled professionals to employ their professionals skills without being asked to formulate strategies for attracting and maintaining new business. If up-or-out is one of several tools that professional service firms use to search efficiently for partners, then the decision of firms not to negotiate a new deal with skilled professionals who turn out not to be well-suited for the partner role is not a riddle that needs to be solved. Associates enter these firms because they hope to become partners. Once they realize that they will not become partners, they also know that their best options are elsewhere.

6. Empirical Patterns Concerning Hours and Promotions

So far, we have discussed up-or-out rules as uniform policies that firms apply rigidly in a firm or industry. Yet, professional service firms differ in how strictly

[7]One could always implement the same outcomes using only two contracts: the outside sector contract we describe above and a single professional contract that covers work loads, pay, retention, and assignment in professional firms. The latter contract would specify all second period outcomes as functions of first-period output signals, and this two contract approach would generate the same allocations and payoffs that we describe in Appendix B. The distinction between firms who dismiss certain workers and firms who offer the same workers contracts that are always rejected is not meaningful in our context.

they adhere to up-or-out policies. Many large firms in law and public accounting have some permanent roles for senior professionals who possess special skill and expertise but are not well matched with the business development tasks that partners perform. Yet, even in large firms, the number of these senior non-partner roles is limited, and most new professionals begin their career expecting to either earn promotion to partner or leave their initial employers.

Nonetheless, the outcomes we observe among those who occupy these senior non-partner roles help us understand why firms ask young professionals to take on heavy work loads as they begin employments path that will most likely end with an up or out decision. Table 2 describes data that we purchased from a firm that researches the private law firm market. The data come from eight annual surveys taken during the period 2007-2014. Some firms appears 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.

Table 2 describes outcomes for lawyers who are between eight and twelve years into their careers. We chose this experience interval because firms make many crucial retention and promotion decisions about progression to partner in this interval. The table presents results for associates, equity partners, non-equity partners, and counsel attorneys. Our reading indicates that different law firms use the titles non-equity partner and counsel to describe different roles. However, in these data, the set of non-equity partners contains some persons who share in firm revenues and likely some who are still being considered for promotion to a full equity partnership. On the other hand, the documentation contains no evidence that counsel attorneys share revenues or are being considered for partnership status.

Because these counsel attorneys are less than 12 years into their careers, it seems reasonable to assume that they are attorneys who recently left an associate position in their current or previous firms and now occupy a senior role off the partnership track. Given this assumption, 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. Associates work fewer hours than either equity or non-equity partners, but these hours may understate the hours of associates who are still fully engaged in up-or-out promotion contests. We have learned through our background research that, when firms tell associates that they are going to dismiss them, firms often reduce their work loads and provide placement services that facilitate their searches for new positions. So, it is noteworthy that counsels bill their time at rate 17 percent greater than associates yet bill 25 percent fewer hours.

These counsel attorney are skilled lawyers. If they were not, they could not bill their time as such high rates. Yet, they work significantly less than the lawyers on the partnership track. 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.

Gicheva [2013] points out that many young professionals work long hours to acquire human capital that allows them to progress in their career, and it seems reasonable to expect that human capital accumulation is an important benefit of long hours among young professionals. However, human capital models cannot explain the patterns in Table 2. All of these lawyers have roughly the same level of experience. If these counsel attorneys were simply lawyers who began their careers with low levels of ambition and chose to work fewer hours all along, we would not expect them to command such high billing rates. It seems more likely that these counsels began as associates and cut back their hours when they left the partnership track.

Finally, these results are not the result of women choosing to work fewer hours in order to spend more time with children. We have created a similar table that contains only male lawyers, and the results are quite similar.

Table 2: Hours Worked, Hourly Rate, and Total Compensation by Job Status

Position	Equity Partner	Non-Equity Partner	Associate	Counsel
Hours Billed	1,645	1,611	1,494	1,128
	(417)	(539)	(592)	(632)
N	3,004	3,139	7,177	559
Hourly Rate	276	285	248	290
	(74)	(92)	(81)	(99)
N	3,012	3,135	6,964	544
Compensation	219,655	188,228	134,219	133,497
	(93,727)	(76,515)	(53,903)	(69,740)
N	3,075	3,192	7,316	575

This table reports job characteristics for lawyers included in the Small Law Firm Economic survey 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. Billable hours is the annual number of billable hours for each attorney. Hourly rate is the typical hourly rate charged by each attorney.

7. Conclusion

The literatures on up-or-out rules and heavy work loads for young professionals are motivated by striking features of the same labor markets. To date, economists have sought to understand these features as separate phenomena. We argue that both practices serve the same purpose. Professional service firms prosper when they identify talented partners who can attract and maintain valuable clients. Thus, the work in professional service firms is performed primarily by partners or by new associates who are participating in strenuous auditions for partner positions.

8. Appendix A

Claim 1: $V(\rho^u,\rho^x) = K_1 + K_2\rho^u + K_3\rho^x$ for some constants K_1, K_2, and K_3.
Proof of Claim 1: Recall that $V(\rho^u,\rho^x)$ satisfies the Bellman equation

$$\begin{aligned} V(\rho^u,\rho^x) &= \max_{\alpha^x,\alpha^u,n} (2-q)\,v^o + qv_0^a(n) + \alpha^x\rho^x\left(v_1^p - v_0^a(n)\right) + \\ &\qquad \alpha^u\rho^u\left(v_1^a - v_0^a(n)\right) + \beta V\left(\rho^u_{+1},\rho^x_{+1}\right) \end{aligned} \tag{8.1}$$

where

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

This equation can be written as

$$V = T(V)$$

where T is an operator defined over the space of bounded functions on the domain $\left\{(\rho^u,\rho^x)\in\mathbb{R}^2_+ : 0 \le \rho^u + \rho^x \le 1\right\}$.

First, we argue that T is a contraction, i.e. for any two functions V_1 and V_2, $|T(V_1) - T(V_2)| < |V_1 - V_2|$. For this, it will be enough to verify Blackwell's sufficient conditions for T to be a contraction:

(1) Monotonicity: If $V_1 \le V_2$ then $T(V_1) \le T(V_2)$
(2) Discounting: there exists some $\beta\in(0,1)$ such that $T(V+a) \le T(V) + \beta a$ for all $a \ge 0$.

Both of these are straightfoward to verify. Define

$$\begin{aligned} \overline{T}(V;\alpha^x,\alpha^u,n) &= (2-q)\,v^o + qv_0^a(n) + \alpha^x\rho^x\left(v_1^p - v_0^a(n)\right) + \\ &\qquad \alpha^u\rho^u\left(v_1^a - v_0^a(n)\right) + \beta V\left(\rho^u_{+1},\rho^x_{+1}\right) \end{aligned}$$

i.e. $\overline{T}(V)$ corresponds to $T(V)$ evaluated not at the vector (α^x,α^u,n) that maximizes the RHS of (8.1) but at an arbitrary vector (α^x,α^u,n). Let $(\widehat{\alpha}^x,\widehat{\alpha}^u,\widehat{n})$ be the vector that maximizes the RHS of (8.1) when $V = V_1$. If $V_2 \ge V_1$, then we then have

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

This establishes that T satisfies 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. To prove this, it will be enough to show that if V is a linear function in (ρ^x,ρ^u), then $T(V)$ must be linear in (ρ^x,ρ^u) as well. The contraction mapping theorem would then imply that there must exist a fixed point within the set of linear functions V. Since the value function corresponds to the unique fixed point $V = T(V)$, it follows that $V(\rho^x,\rho^u)$ belongs to the set of linear functions V.

Suppose that V is linear in ρ^u and ρ^x, i.e.

$$V(\rho^u, \rho^x) = K_1' + K_2'\rho^u + K_3'\rho^x \tag{8.2}$$

Since $v_0^a(n)$ is concave in n, the RHS of (8.1) is concave in n. Hence, for any value of α^x and α^u, the first order condition with respect to n represents a necessary and sufficient condition for optimality:

$$(q - \rho^u\alpha^u - \rho^x\alpha^x)\frac{dv_0^a}{dn} + (q - \rho^u\alpha^u - \rho^x\alpha^x)\beta\left(\pi\frac{x}{\varepsilon}K_3' - \frac{x}{\varepsilon}K_2'\right) = 0$$

Dividing through by $(q - \rho^u\alpha^u - \rho^x\alpha^x)$, we are left with a first order condition

$$c'(n) = (1 + \pi x) + \beta\left(\pi\frac{x}{\varepsilon}K_3' - \frac{x}{\varepsilon}K_2'\right)$$

It follows that the n which solves (8.1) is independent of ρ^u and ρ^x, and depends only on 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:

$$\alpha^u = \begin{cases} 1 & \text{if } v_1^a - v_0^a - \beta\left[\pi\frac{x}{\varepsilon}nK_3' + \left(1 - \frac{x}{\varepsilon}n\right)K_2'\right] \geq 0 \\ 0 & \text{if } v_1^a - v_0^a - \beta\left[\pi\frac{x}{\varepsilon}nK_3' + \left(1 - \frac{x}{\varepsilon}n\right)K_2'\right] < 0 \end{cases} \tag{8.3}$$

$$\alpha^x = \begin{cases} 1 & \text{if } v_1^p - v_0^a - \beta\left[\pi\frac{x}{\varepsilon}nK_3' + \left(1 - \frac{x}{\varepsilon}n\right)K_2'\right] \geq 0 \\ 0 & \text{if } v_1^p - v_0^a - \beta\left[\pi\frac{x}{\varepsilon}nK_3' + \left(1 - \frac{x}{\varepsilon}n\right)K_2'\right] < 0 \end{cases} \tag{8.4}$$

Note that we can set α^x and α^u to 1 when the expression on the RHS is exactly equal to 0 without loss of generality, since when this expression is equal to 0 any value of α^x or α^u yields the same value for the objective function. Since n and v_0^a are independent of ρ^x and ρ^u, it follows that α^x and α^u are as well, i.e. α^x and α^u can also be expressed as functions of K_2' and K_3'. Hence, if V is given by (8.2), then

$$T(V) = K_1'' + K_2''\rho^u + K_3''\rho^x$$

where

$$\begin{aligned} K_1'' &= v^o + v_0^a(n) + \beta q\left[\pi\frac{x}{\varepsilon}nK_3' + \left(1 - \frac{x}{\varepsilon}n\right)K_2'\right] + \beta K_1' \\ K_2'' &= \alpha^u(v_1^a - v_0^a) - \beta\alpha^u\left[\pi\frac{x}{\varepsilon}nK_3' + \left(1 - \frac{x}{\varepsilon}n\right)K_2'\right] \\ K_3'' &= \alpha^x(v_1^p - v_0^a) - \beta\alpha^x\left[\pi\frac{x}{\varepsilon}nK_3' + \left(1 - \frac{x}{\varepsilon}n\right)K_2'\right] \end{aligned}$$

and $n = n(K_2', K_3')$, $\alpha^u = \alpha^u(K_2', K_3')$, and $\alpha^x = \alpha^x(K_2', K_3')$. That is, if V is linear with coefficients K_1', K_2', and K_3', then $T(V)$ will also be linear with coefficients K_1'', K_2'', and K_3''. The claim thus follows.

Claim 2: $\widehat{\alpha}^x = 1$.

Proof of Claim 2: Define $m \equiv \frac{x}{\varepsilon}n$. Since $\lim_{n\to\overline{n}} c'(n) = \infty$, we know the optimal $n < \overline{n}$. It follows that at the optimal n, the expression $m < 1$, since

$$m \equiv \tfrac{x}{\varepsilon}n < \tfrac{x}{\varepsilon}\overline{n} \leq 1$$

Since $V(\rho^u, \rho^x)$ is linear, we can write it as

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

Matching coefficients, we can deduce that K_1, K_2, and K_3 must satisfy the following equations for the Bellman equation to hold:

$$K_1 = v^o + v_0^a + \beta q\left[\pi\frac{x}{\varepsilon}nK_3 + \left(1-\frac{x}{\varepsilon}n\right)K_2\right] + \beta K_1 \quad (8.5)$$

$$K_2 = \alpha^u\left(v_1^a - v_0^a\right) - \beta\alpha^u\left(\pi m K_3 + (1-m)K_2\right) \quad (8.6)$$

$$K_3 = \alpha^x\left(v_1^p - v_0^a\right) - \beta\alpha^x\left(\pi m K_3 + (1-m)K_2\right) \quad (8.7)$$

Using these equations to solve for K_2 and K_3 yields

$$K_2 = \frac{\alpha^u\left(v_1^a - v_0^a\right) - \beta\alpha^u\alpha^x\pi m\left(v_1^p - v_1^a\right)}{1+\beta\alpha^u(1-m)+\beta\alpha^x\pi m} \quad (8.8)$$

$$K_3 = \frac{\alpha^x\left(v_1^p - v_0^a\right) + \beta\alpha^u\alpha^x(1-m)\left(v_1^p - v_1^a\right)}{1+\beta\alpha^u(1-m)+\beta\alpha^x\pi m} \quad (8.9)$$

The first order condition for α^x, which is the analog of (8.4) but with $K_2' = K_2$ and $K_3 = K_3'$, implies that $\alpha^x = 1$ will be optimal whenever

$$v_1^p - v_0^a \geq \beta\left(\pi m K_3 + (1-m)K_2\right)$$

Hence, it will suffice to show that both K_2 and K_3 are bounded above by $v_1^p - v_0^a$, since as long as $m \in [0,1]$ we would have

$$\beta\left(\pi m K_3 + (1-m)K_2\right) \leq \beta\left(v_1^p - v_0^a\right) < v_1^p - v_0^a$$

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

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

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

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

It follows that $\widehat{\alpha}^x = 1$ is optimal.

Lemma 1: $\pi K_3 > K_2$

Proof of Lemma 1: Observe that $K_1 - (1-\beta)^{-1}(2-q)v^o$ represents the surplus generated in the professional sector from 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 acting optimally 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

$$(1-\beta)K_1 > (2-q)v^o + qv_0^a + \beta qK_2 \tag{8.10}$$

From (8.5) in the proof of Claim 2, we know that K_1 must satisfy the following equality:

$$(1-\beta)K_1 = (2-q)v^o + qv_0^a + \beta q\pi mK_3 + \beta q(1-m)K_2$$

Rearranging this equation implies

$$\beta m(\pi K_3 - K_2) = (1-\beta)K_1 - (2-q)v^o - qv_0^a - \beta qK_2 \tag{8.11}$$

The RHS of (8.11) is strictly positive given (8.10). Since $m > 0$ given it will always be optimal to have inexperienced workers put in some effort, Hence, $\beta m(\pi K_3 - K_2) > 0$. It follows that $\pi K_3 - K_2 > 0$.

Proposition 1: The optimal work load for new associates, $\widehat{n}$, exceeds the static optimum implied by the expected per period output of new associates.

Proof of Proposition 1: The first order condition for $\widehat{n}$ is given by

$$c'(\widehat{n}) = (1+\pi x) + \beta\frac{x}{\varepsilon}(\pi K_3 - K_2) \tag{8.12}$$

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

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

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

Proposition 2: $\widehat{\alpha}^u = 0$ if $v_1^a - v_0^a(n) < \beta\pi\frac{x}{\varepsilon}\widehat{n}(v_1^p - v_1^a)$

Proof of Proposition 2: The first-order condition for α^u implies that

$$\widehat{\alpha}^u = \begin{cases} 1 & \text{if } v_1^a - v_0^a - \beta[\pi mK_3 + (1-m)K_2] > 0 \\ [0,1] & \text{if } v_1^a - v_0^a - \beta[\pi mK_3 + (1-m)K_2] = 0 \\ 0 & \text{if } v_1^a - v_0^a - \beta[\pi mK_3 + (1-m)K_2] < 0 \end{cases} \tag{8.13}$$

where $m = \frac{x}{\varepsilon}\widehat{n}$. Substituting in for K_2 and K_3 from the proof of Claim 2, we have

$$\pi m K_3 + (1-m) K_2 = \frac{\pi m \left(v_1^p - v_0^a\right) + \alpha^u (1-m) \left(v_1^a - v_0^a\right)}{1 + \beta\alpha^u (1-m) + \beta\pi m}$$

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

$$v_1^a - v_0^a \geq \beta \frac{\pi m \left(v_1^p - v_0^a\right) + \alpha^u (1-m) \left(v_1^a - v_0^a\right)}{1 + \beta\alpha^u (1-m) + \beta\pi m}$$

which implies

$$\begin{aligned} \left(1 + \beta\alpha^u (1-m) + \beta\pi m\right) \left(v_1^a - v_0^a\right) &\geq \beta\pi m \left(v_1^p - v_0^a\right) + \beta\alpha^u (1-m) \left(v_1^a - v_0^a\right) \\ \left(1 + \beta\pi m\right) \left(v_1^a - v_0^a\right) &\geq \beta\pi m \left(v_1^p - v_0^a\right) \\ \left(v_1^a - v_0^a\right) &\geq \beta\pi m \left(v_1^p - v_1^a\right) \end{aligned} \tag{8.14}$$

Note that this implies the optimal α^u will be given by

$$\alpha^u = \begin{cases} 1 & \text{if } v_0^a(n) < v_1^a - \beta\pi\frac{x}{\varepsilon} n \left(v_1^p - v_1^a\right) \\ [0,1] & \text{if } v_0^a(n) = v_1^a - \beta\pi\frac{x}{\varepsilon} n \left(v_1^p - v_1^a\right) \\ 0 & \text{if } v_0^a(n) > v_1^a - \beta\pi\frac{x}{\varepsilon} n \left(v_1^p - v_1^a\right) \end{cases} \tag{8.15}$$

Since $v_0^a(n) + \beta\pi\frac{x}{\varepsilon} n \left(v_1^p - v_1^a\right)$ is concave in n, it can equal the constant v_1^a for at most two values of n. That is, there exist two values $n_1 < n_2$ such that

$$\alpha^u \in \begin{cases} 0 & \text{if } n \in (n_1, n_2) \\ [0,1] & \text{if } n = n_1 \text{ or } n = n_2 \\ 1 & \text{if } n < n_1 \text{ or } n > n_2 \end{cases} \tag{8.16}$$

We will use this characterization for α^u in the proof of Propositions 3 and 4.

Proposition 3: The optimal work load for new associates, $\widehat{n}$, is increasing in z^p and non-increasing in z^a.

Proposition 4: The optimal $\widehat{\alpha}^u$ is non-increasing in z^p and non-decreasing in z^a.

Proof of Propositions 3 and 4: We first consider the problem of choosing a value for n to maximize the utility of the planner given a value of α^u. Denote this value by $n^*(\alpha^u)$. We show that $n^*(\alpha^u)$ is increasing in z^p and non-decreasing in z^a. We then argue that $\widehat{\alpha}^u$, the value of α^u which maximizes the utility of the planner together with n, is non-increasing in z^p and non-decreasing in z^a. Finally, we use these two results to conclude that when we choose both α^u and n optimally, the optimal $\widehat{n}$ is increasing in z^p and non-decreasing in z^a.

Since $n^*(\alpha^u)$ maximizes the value to the planner for a given value of α^u, it must satisfy the first order necessary condition

$$\left.\frac{\partial K_1}{\partial n}\right|_{n=n^*} = 0 \tag{8.17}$$

To see how $n^*(\alpha^u)$ varies with z^p, we can look at how it varies with v_1^p given the latter is montonically increasing in z^p. Totally differentiate (8.17) to obtain

$$\frac{dn^*}{dv_1^p} = -\left.\frac{\partial^2 K_1/\partial v_1^p \partial n}{\partial^2 K_1/\partial n^2}\right|_{n=n^*}$$

Using the expressions for (8.8) and (8.9), one can show[8] that

$$\frac{\partial^2 K_1}{\partial v_1^p \partial n} = \frac{\beta\pi\frac{x}{\varepsilon}\left(1+\alpha^u\beta\right)}{\left(1+\beta\alpha^u\left(1-m\right)+\beta\pi m\right)^2}$$

This equality holds for all values of n. The expression for $\partial^2 K_1/\partial n^2$ is more complicated. However, the value n^* must satisfy the first order necessary condition (8.12) when K_2 and K_3 are equal to $K_2\left(n^*\right)$ and $K_3\left(n^*\right)$, i.e.

$$c'\left(n^*\right) = \left(1+\pi x\right) + \beta\frac{x}{\varepsilon}\left[\pi K_3\left(n^*\right) - K_2\left(n^*\right)\right]$$

Evaluated at this condition, we can deduce that

$$\left.\frac{\partial^2 K_1}{\partial^2 n}\right|_{n=n^*} = \frac{-c''\left(n^*\right)}{1+\beta\alpha^u\left(1-m\right)+\beta\pi m}$$

Taking the ratio of the two expressions reveals that

$$\frac{dn^*}{dv_1^p} = \frac{\beta\pi\frac{x}{\varepsilon}\left(1+\alpha^u\beta\right)}{\left(1+\beta\alpha^u\left(1-m\right)+\beta\pi m\right)c''\left(n^*\right)} > 0$$

In other words, increasing v_1^p will induce the planner to choose a higher n^*. By an analogous argument,

$$\frac{dn^*}{\partial v_1^a} = -\left.\frac{\partial^2 K_1/\partial v_1^a \partial n}{\partial^2 K_1/\partial n^2}\right|_{n=n^*}$$

In this case,

$$\frac{\partial^2 K_1}{\partial v_1^a \partial n} = -\frac{\alpha^u\beta\pi\frac{x}{\varepsilon}\left(1+\pi\beta\right)}{\left(1+\beta\alpha^u\left(1-m\right)+\beta\pi m\right)^2}$$

and so

$$\frac{dn^*}{dv_1^a} = -\frac{\alpha^u\beta\frac{x}{\varepsilon}\left(1+\pi\beta\right)}{\left(1+\beta\alpha^u\left(1-m\right)+\beta\pi m\right)c''\left(n^*\right)} \leq 0$$

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

Next, we want to argue that the optimal $\widehat{\alpha}^u$ can only change between 0 and 1 exactly once as we vary either v_1^p or v_1^a. Recall that the derivative of $n^*\left(\alpha^u\right)$, the value of n that maximizes the value of the planner for a given α^u, with respect to v_1^p and v_1^a are given by

$$\begin{aligned}
\frac{dn^*}{dv_1^p} &= \frac{\beta\pi\frac{x}{\varepsilon}\left(1+\alpha^u\beta\right)}{\left(1+\beta\alpha^u\left(1-m\right)+\beta\pi m\right)c''\left(n^*\right)} \\
\frac{dn^*}{dv_1^a} &= -\frac{\alpha^u\beta\frac{x}{\varepsilon}\left(1+\pi\beta\right)}{\left(1+\beta\alpha^u\left(1-m\right)+\beta\pi m\right)c''\left(n^*\right)}
\end{aligned}$$

respectively. We now argue that this implies there can be at most one value of v_1^p and one value of v_1^a for which $n^*\left(0\right) = n^*\left(1\right)$. In the case of v_1^a, this is immediate: $n^*\left(0\right)$ does not vary with v_1^a while $n^*\left(1\right)$ is decreasing with v_1^a, so they can equal at most once. In the case of v_1^p, observe that the derivative of n^* with respect to v_1^p is increasing in α^u, i.e.

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

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

Hence, 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 positive. This implies there can be at most one value v_1^p for which $n^*(1; v_1^p) - n^*(0; v_1^p) = 0$.

From the optimality condition (8.16) for α^u, we know that the optimal rule will either set α^u to 0 or to 1, unless the optimal $\widehat{n} \in \{n_1, n_2\}$, in which case any $\alpha^u \in [0,1]$. Hence, as we increase v_1^p (alternatively, v_1^a), the optimal $\widehat{\alpha}^u$ will not change except at the one value of v_1^p for which $n^*(1; v_1^p) = n^*(0; v_1^p)$ in which any $\widehat{\alpha}^u$ is optimal (alternatively, except at the one value of v_1^a for $n^*(1; v_1^a) = n^*(0; v_1^a)$ in which any $\widehat{\alpha}^u$ is optimal). Hence, for all values of v_1^p other than the one threshold (alternatively, at all values of v_1^a other than the one threshold), $\frac{d\widehat{n}}{dv_1^p}$ and $\frac{d\widehat{n}}{dv_1^a}$ are equal to $\frac{dn^*}{dv_1^p}$ and $\frac{dn^*}{dv_1^a}$ when we evaluate the latter derivative at $n^*(\widehat{\alpha}^u)$, i.e. at the effort level that maximizes the planner's utility when we set α^u to the level it would assume at the unconstrained optimum. For the threshold v_1^p, since any $\alpha^u \in [0,1]$ is optimal, we can use the derivative evaluated at $n^*(\widehat{\alpha}^u)$ depending on whether we are taking the derivative from the left or the right. Hence, we can conclude that $\widehat{n}$ is increasing in v_1^p and non-increasing in v_1^a, since $\frac{dn^*}{dv_1^p} > 0$ and $\frac{dn^*}{dv_1^a} \leq 0$.

Finally, since we know α^u can switch 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 $\widehat{\alpha}^u$. In the limit as $z^a \to 1$, it be optimal to set $\alpha^u = 0$ and employ a young worker who has some option value than an experienced worker 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 $\alpha^u = 1$.

Proposition 5: $\widehat{q}$ is increasing in z^p.

Proof of Proposition 3: From the first order condition for the planner's problem, we have

$$\begin{aligned} \kappa(q) &= v_0^a - v^o + \beta\left[\pi\frac{x}{\varepsilon}n\frac{\partial V}{\partial \rho_t^x} + \left(1 - \frac{x}{\varepsilon}n\right)\frac{\partial V}{\partial \rho_t^u}\right] \\ &= v_0^a - v^o + \beta\left[\pi m K_3 + (1-m)K_2\right] \qquad (8.18) \end{aligned}$$

From Proposition 2, we know that α^u is nondecreasing in v_1^p, and there exists a single value of v_1^p for which any $\alpha^u \in [0,1]$ is optimal, including both 0 and 1. Hence, without loss of generality, we can treat the optimal α^u as fixed as we change

v_1^p by a sufficiently small amount. If we held n fixed, the expression

$$\pi m K_3 + (1-m) K_2 = \frac{\pi m \left(v_1^p - v_0^a\right) + \alpha^u (1-m) \left(v_1^a - v_0^a\right)}{1 + \beta \alpha^u (1-m) + \beta \pi m}$$

is strictly increasing in v_1^p. Next, since n must maximize K_1, we know that

$$\frac{d}{dn} \left[v_0^a + \beta \left(\pi m K_1 + \beta (1-m) K_2\right)\right] = 0$$

Hence, at the optimum, changing n will have no effect on the RHS of (8.18). Thus, at the optimal allocation, the RHS of (8.18) must be increasing in v_1^p. Since $\kappa(q)$ is increasing in q, it follows that the optimal q will be higher as well.

9. Appendix B

Forthcoming

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.

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

Suman Ghosh and Michael Waldman. Standard promotion practices versus up-or-out contracts. *The RAND Journal of Economics*, 41(2):201–325, 2010.

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

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.

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.

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