

ALLOCATING EFFORT AND TALENT IN PROFESSIONAL LABOR MARKETS

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

ABSTRACT. In many professional service firms, new associates work long hours while competing in up-or-out promotion contests. Our model explains why young professionals who face up-or-out promotion rules also often work longer hours than other young professionals. 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: October 31, 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 address retention and 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 law, consulting, investment banking and several other professions often work much longer hours

than most white collar workers who have similar levels of education. The theoretical literature that addresses this pattern highlights reasons that new associates in professional labor markets may work more than the efficient number of hours.

? focus on the fact that law firms are typically organized as partnerships. They view law firms as teams and assume that, while team output is observed, individual output is not observed. 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.

The key assumptions are that workers 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 ? 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 ?, hours distortions are the equilibrium mechanism that sorts heterogeneous workers to heterogeneous teams. Yet, the model only addresses the initial sorting of new lawyers to law firms. It does not address re-assignments within or between firms, and thus, ? do not address specific promotion or retention rules like up-or-out. In their separating equilibrium, no one leaves law as a profession, no one changes law firms, and all associates become partners.

? provides a different reason that young professionals may work excessive hours. In his model, workers are heterogeneous with respect to their abilities, but workers do not have private information about their abilities. All market participants are symmetrically informed about the abilities of all workers and learn about all workers at the same rate by observing public output signals. Output is not contractible, so firms pay workers ex ante based on their reputation.

Although output signals are public, workers have private information about their effort levels, and ? shows that young workers may work more than efficient levels to manipulate market beliefs about their abilities. In equilibrium, firms know that workers engage in this manipulation 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

We note above that, in the ? model, all associates make partner in the firm they choose. Holmstrom's model does not address promotion or retention at all. There is only one job in the model, and all workers earn their expected marginal product given their production in previous periods.

¹In 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.

Just as the literature on excess hours among young professionals does not address promotion and retention directly, the literature on up-or-out 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 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 ? who argue that up-or-out allows firms to induce workers to make investments in firm-specific skills. ? argues that up-or-out rules are not always needed to solve the double moral hazard problem that ? 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 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.

? extends the logic of ? 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 ? 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.

? extend ? 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. These are settings where firms demand a mechanism that allows them to commit to efficient promotion decisions. Here, the ex post surplus generated by efficient promotions is relatively low. So, firms face relatively strong incentives to renege on contingent promises concerning raises and promotions.²

²? also model worker effort. However, in contrast to our model, worker actions are hidden in their model. As in ?, workers may expend effort to 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.

A separate literature on up-or-out promotion rules links up-or-out rules to optimal screening procedures. ? 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.

? 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.

Both ? and ? describe 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 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, we describe up-or-out equilibria such that the static surplus generated by newly hired professionals is strictly less than the static surplus that could have been generated by more experienced professionals that they are replacing. Our up-or-out result speaks directly to why up-or-out firms do not make retention offers to experienced professionals who perform well in their current roles but are not well suited to the job of partner.

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 rules out 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 ? and ? since

³? 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

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 demonstrate that optimal work loads for new professionals interact with up-or-out rules as components of a set of optimal screening procedures. Further, 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. In the professional sector, output does vary with ability, given effort, and the return to ability is greatest in the partner position.

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 where all firms adopt the same personnel policies, and the competitive equilibrium in this market produces the allocations that solve our planners 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 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 instead, the planner 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

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.

about associate talent levels and therefore permit the identification and promotion of more new partners.

We begin by describing the production environment and the learning technology. Next, we derive optimal policies concerning effort levels and job assignments. Finally, we show that simple decentralized mechanisms implement these optimal policies.

2.1. Economic Environment. 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 effort, and $c(n)$ is the disutility of effort. Below, we describe a planner's problem in which the planner assigns workers to jobs and effort levels. These effort assignments are equivalent to task assignments or work loads. Here, we assume that all workers apply the same focus and intensity to their work. Thus, worker A expends twice as much effort as worker B if and only if worker A completes a set of tasks that requires twice as much time input. We assume $c(0) = 0$, $c'(0) = 0$. Further, $\lim_{n \rightarrow \bar{n}} c'(n) = \infty$, $c''(n) > 0 \quad \forall n \in [0, \bar{n}]$.

Let θ denote worker ability, which is either high or low, i.e. $\theta \in \{0, x\}$, with $x > 0$. 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

$$(2.1) \quad y_0^a = (1 + \theta)n + \epsilon$$

The production function for experienced associates is

$$(2.2) \quad y_1^a = z^a(1 + \theta)n + \epsilon$$

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

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

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 function 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 this 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

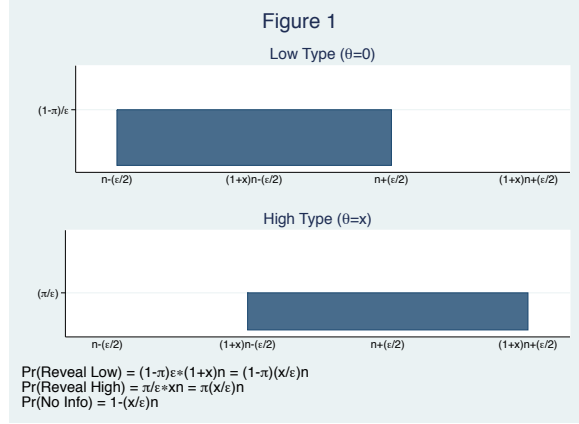
$$(2.4) \quad z^a < w^o < z^a(1 + \pi x)$$

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 .

We follow Pries (2004) and assume that the production shocks are uniformly distributed, $\epsilon \sim U[\frac{-\epsilon}{2}, \frac{\epsilon}{2}]$. 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 how learning takes place. The two panels in the figure present two joint densities. A given area under the density in the top panel equals the joint probability that a new associate with a work load of n is both 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 $(1+x)n - \frac{\epsilon}{2}$, 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 $n + \frac{\epsilon}{2}$, 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 $\frac{\pi}{\epsilon}$ while the joint density in the top panel is $\frac{1-\pi}{\epsilon}$. Thus, Bayes' rule implies that

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

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, $[(1+x)n - \frac{\epsilon}{2}, n + \frac{\epsilon}{2}]$, is $\epsilon - xn$. Multiply this length by the density of the production shock, $\frac{1}{\epsilon}$, to get, $1 - \frac{x}{\epsilon}n$, which is the probability that the output signal reveals no information about associate ability. It follows immediately that $\pi \frac{x}{\epsilon}n$ is the probability that the output signal reveals that an associate has high ability, and $(1-\pi) \frac{x}{\epsilon}n$ is the probability that the signal reveals that an associate has low ability. We assume that $\frac{x}{\epsilon}n < 1$ 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 analyses 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	<i>Partner</i>	?

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. To see this, first consider the top right corner of Table 1. These are new workers who by assumption are of uncertain ability. Since low ability workers produce output equal to $-\infty$ if they work as partners, the planner either assigns new workers to the outside sector or to associate positions in the professional sector. 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.

Next, consider the bottom left hand corner. These are workers who 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 who 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 represent 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 the 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. We know that, as long as the planner retains some experienced professionals in the professional sector, the employment restriction on the professional sector must bind, and the planner must assign some of these new

workers to the outside sector. 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, it takes little work to establish 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 stay 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 surplus created by a new associate who takes on a work load of n .
- v_1^a - the surplus created by an experienced professional of uncertain ability who works as an associate
- v_1^p - the 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, all partners must have high ability and all experienced associates must possess uncertain ability.

Next, 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 limit on professional employment, q , always binds. This implies that a mass of $2 - q$ of workers works in the outside sector while the mass q of workers in the professional sector is divided among new associates and two types of experienced professionals. Since the planner assigns all workers with known low-ability or no professional experience to the outside sector, the planner's assignment decisions concerning the stocks of experienced professionals who do not possess known low ability pin down how the mass of new workers is divided 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

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

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

$$\begin{aligned}
 V(\rho^u, \rho^x) &= \max_{n, \alpha^u, \alpha^x} (2 - q)v^o + qv_0^a(n) + \alpha^u \rho^u (v_1^a - v_0^a(n)) + \alpha^x \rho^x (v_1^p - v_0^a(n)) + \\
 (3.1) \quad &\beta V \left([q - \alpha^u \rho^u - \alpha^x \rho^x] \left[1 - \frac{x}{\varepsilon} n \right], [q - \alpha^u \rho^u - \alpha^x \rho^x] \pi \frac{x}{\varepsilon} n \right)
 \end{aligned}$$

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. Output this period 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 $q\pi\frac{x}{\varepsilon}n$.

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 $\alpha^x = 1$.

3.2. Promotion to Partner. Appendix A demonstrates that $V(\rho^u, \rho^x)$ defined in equation 3.1 is a contraction mapping. Given this result, it is easy to establish the following claim:

Claim 1. $V(\rho^\theta, \rho^u)$ is linear.

Proof of Claim 1: Equation 3.1 can be expressed as

$$V = T(V)$$

where T is an operator defined over the space of all bounded functions that are defined over the domain $\{0 \leq \rho^u + \rho^\theta \leq 1\}$.

Assume that $V(\rho^u, \rho^x) = K_1 + K_2\rho^u + K_3\rho^x$. This implies that

$$\begin{aligned} K_1 + K_2\rho^u + K_3\rho^x &= \max_{n, \alpha^u, \alpha^x} (2 - q)v^o + qv_0^a + \\ &\quad \alpha^u\rho^u(v_1^a - v_0^a(n)) + \alpha^x\rho^x(v_1^p - v_0^a(n)) + \\ (3.2)\beta [K_1 + (q - \alpha^u\rho^u - \alpha^x\rho^x)(1 - \frac{x}{\varepsilon}n)K_2 + (q - \alpha^u\rho^u - \alpha^x\rho^x)\pi\frac{x}{\varepsilon}nK_3] \end{aligned}$$

The first order conditions for this problem are

$$(3.3) \quad c'(n) = (1 + \pi\theta) + \beta\frac{x}{\varepsilon}(\pi K_3 - K_2)$$

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

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

Note that none of these conditions involve ρ^u or ρ^x . The optimal policies are independent of the current stocks with one exception. The mass of new associates who enter the professional sector as opposed to the outside sector each period equals one minus the mass of experienced professionals retained from the current stocks ρ^u and ρ^x . However, the retention decisions concerning experienced professionals and the work loads assigned to new associates are not functions of these stocks. Let $\{\hat{n}, \hat{\alpha}^u, \hat{\alpha}^x\}$ denote the solutions to these first order conditions. If $T(V)$ is the contraction defined in equation 3.2, it is straightforward to show that

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

where

$$\begin{aligned} \bar{K}_3 &= \hat{\alpha}^x (v_1^x - v_0^a(\hat{n})) - \beta \hat{\alpha}^x \left[\left(1 - \frac{x}{\varepsilon} \hat{n}\right) K_2 + \pi \frac{x}{\varepsilon} \hat{n} K_3 \right] \\ \bar{K}_2 &= \hat{\alpha}^u (v_1^u - v_0^a(\hat{n})) - \beta \hat{\alpha}^u \left[\left(1 - \frac{x}{\varepsilon} \hat{n}\right) K_2 + \pi \frac{x}{\varepsilon} \hat{n} K_3 \right] \\ \bar{K}_1 &= v^0 + v_0^a(\hat{n}) + \beta \left[K_1 + q \left(1 - \frac{x}{\varepsilon} \hat{n}\right) K_2 + q \pi \frac{x}{\varepsilon} \hat{n} K_3 \right] \end{aligned}$$

Because $\{\hat{n}, \hat{\alpha}^u, \hat{\alpha}^x\}$ are functions of K_2 and K_3 , \bar{K}_1 , \bar{K}_2 , and \bar{K}_3 are not simple linear combinations of K_1 , K_2 , and K_3 . However, it is still the case that, given a linear $V(\rho^u, \rho^x)$, the function $T(V)$ is also linear. Since $T(V)$ is linear when V is linear, we know that there is exists a linear $V(\rho^u, \rho^x)$ that satisfies equation 3.1. Given the contraction mapping theorem, we also know that a unique value function solution satisfies equation 3.1. Thus, $V(\rho^u, \rho^x)$ must be linear.

The linearity of $V(\rho^u, \rho^x)$ implies that the planner earns constant returns from two stocks of experienced professionals, which implies that the planner never has an incentive to replace 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. This reasoning is the basis for our second claim.

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

Proof of Claim 2: See Appendix A.

The planner always retains professionals with know high-ability. 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 must make 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 n , the work load for new associates, and the planner must choose $\hat{\alpha}^u$, which determines whether or not experienced associates of uncertain ability remain in the professional sector. Here, we demonstrate how these choices are related.

If we return to equation 3.2 and impose our result $\hat{\alpha}^x = 1$, we can derive the following expressions for K_2 and K_3 .

$$K_2 = \frac{\hat{\alpha}^u(v_1^a - v_o^a(\hat{n}) + \beta\hat{\alpha}^u\pi\frac{x}{\varepsilon}\hat{n}(v_1^x - v_1^a)}{1 + \beta\hat{\alpha}^u(1 - \frac{x}{\varepsilon}\hat{n}) + \beta\pi\frac{x}{\varepsilon}\hat{n}}$$

$$K_3 = \frac{v_1^x - v_o^a(\hat{n}) + \beta\hat{\alpha}^u(1 - \frac{x}{\varepsilon}\hat{n})(v_1^x - v_1^a)}{1 + \beta\hat{\alpha}^u(1 - \frac{x}{\varepsilon}\hat{n}) + \beta\pi\frac{x}{\varepsilon}\hat{n}}$$

Note that K_3 is the value of having one more experienced associate who has known high ability. Since the employment constraint in the professional sector is binding, this is the value generated by replacing a new associate with an experienced high-ability partner. Likewise, K_2 is the value created by replacing a new associate with an experienced associate of unknown ability.

Before proceeding to our main results, we state the following lemma concerning K_3 and K_2 .

Lemma 1. $\pi K_3 > K_2$

Appendix A contains a simple proof of this lemma, but it is useful to consider why this condition must hold. Consider what the planner would be willing to pay a third party in order to learn the true ability of a given experienced associate with unknown ability. For the planner, the value of this third party information would be $[\pi K_3 + (1 - \pi)0] - K_2$. With probability π , the third party would report that the experienced associate has high ability, and the planner would promote her to partner. With probability $(1 - \pi)$, the third party would report low ability, and the planner would assign the experienced associate to the outside sector. Because the constraint on professional employment binds in this model, the planner would also move an inexperienced outside worker into a new associate position. Thus, relative to the value the planner places on new associates, the planner attaches values K_3 and 0 to the outcomes that would be generated by the two possible reports. Since K_2 is the value of a experienced associate with unknown ability and no report, the expected value of the report is $\pi K_3 - K_2$. Thus, our lemma states that our planner is willing to pay for information about the ability of experienced professionals. This condition should hold in any model where learning creates surplus by improving job assignments.

With this result in hand, recall equation 3.3. Two results concerning new associate effort are immediate.

Claim 3. The optimal work load for new associates, \hat{n} , exceeds the static optimum implied by the expected per period output of new associates.

Again, the proof in Appendix A is straightforward. The optimal effort condition, equation 3.3, states that the planner chooses a work load that equates the marginal cost of effort with two returns from effort. New associate effort produces not only current period output but also information about worker ability that the planner uses next period.

The parameters z^p and z^a influence the marginal value of identifying a high-ability associate. Assume that an associate produces a first period output signal that reveals she has high ability. The difference between her productivity given promotion to partner and her productivity if she remains an associate is $[z^p - z^a](1 + x)$. Thus, it is easy to show that the value of the information generated by first period output signals is increasing in z^p holding z^a fixed and decreasing in z^a holding z^p fixed. This observation helps us prove the following result concerning work loads for new associates

Proposition 1. *The optimal work load for new associates, \hat{n} , is increasing in z^p and non-decreasing in z^a .*

Appendix A contains the proof. The probability that a new associate’s output signal reveals her true ability increases with 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 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 .

Our model shows that heavy work loads for new associates are part of the solution to a learning and assignment problem. The planner is able to identify more partners next period if new associates perform more than the statically optimal level of work this period.

Up-or-out promotion rules serve a similar purpose. When the planner assigns an experienced associate of uncertain ability to work in the outside sector and replaces her with a new associate, the planner is forfeiting the productivity gains from associate experience in order to begin a search for a new partner one period sooner.

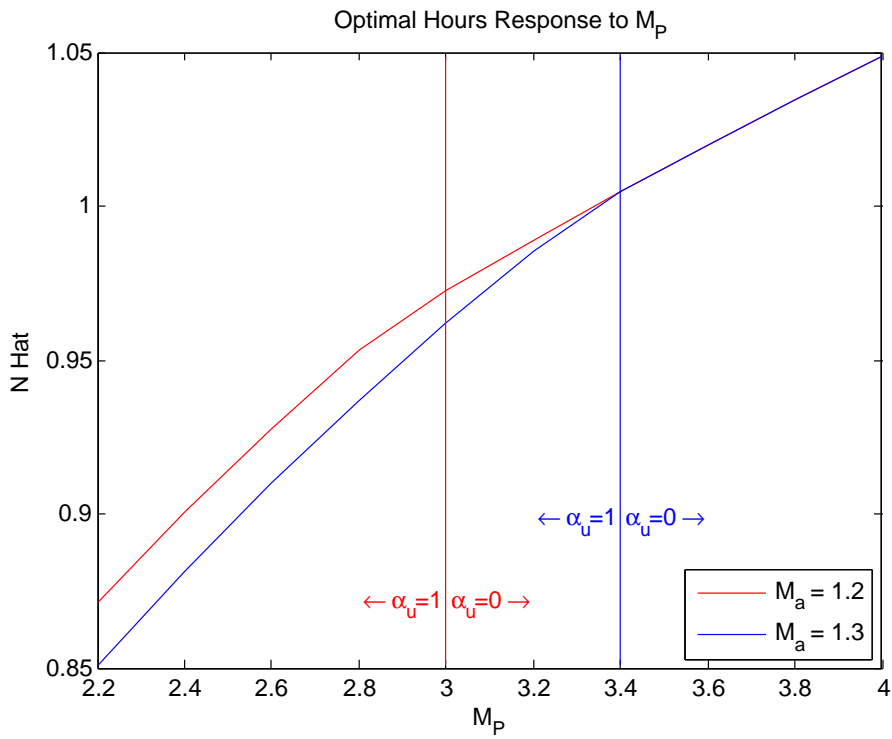
Heavy work loads and up-or-out promotion rules arise when the planner is willing to pay the extra costs generated by these policies in order to increase the number of partners that he discovers. This insight underlies our second proposition.

Proposition 2. *$\hat{\alpha}^u$ is non-increasing in z^p and non-decreasing in z^a . Further, for any combination of $c(n)$, x , π , and ε that satisfy our assumptions above, there exist parameter values for z^p , z^a , and w^0 such that $\hat{\alpha}^u = 0$.*

The policy $\hat{\alpha}^u = 0$ is an up-or-out promotion rule. The planner promotes experienced professionals with known high ability to the partner position and assigns all other experienced professionals to the outside sector.

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} increase 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 .



Whenever the output of a new associate reveals that she has low ability, the planner assigns her to the outside sector and replaces her with a new associate. Further, if the environment is such that the planner follows an up-or-out rule, he also assigns experienced associates of uncertain ability to the outside sector and replaces them with new associates. In contrast to replacements of low-ability professionals, these replacements reduce professional sector output in the current period since experienced associates of uncertain ability are more productive, in expected value, than new associates. Nonetheless, the replacements mandated by up-or-out rules are optimal. Each period, the planner must balance the value of current professional sector output against the value of information that he can use to boost future output. In up-or-out settings, the value of information dominates these calculations.

Both the heavy work loads that the planner assigns to new associates and the up-or-out policies that the planner imposes in some environments make no sense as static allocation rules. Yet, both policies are optimal because the planner is not just producing output for the current period. He is also conducting a search for talent, and the results of this search impact future output.

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 that enter these firms. Our approach is built on the premise that the only way that these elite firms can learn about these talent differences is to have new associates 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 discover the most talented professionals, and this discovery process facilitates efficient matches between task assignments and professional talent levels.

? points out that strict adherence to up-or-out rules became less common in law firms during the 1980s and 1990s. She argues that 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 phrased these claims in terms of the language of our model, ? 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.⁴ Nonetheless, many new associates in elite professional service firms still begin their careers expecting that, within the coming decade, they will most likely either move up to partner or out to another employer.

4.1. Endogenous Sector Size. In our analysis above, we assume that employment in the professional sector is capped at one, and we restrict our attention to parameter vectors such that this constraint is binding. This approach simplifies our presentation, but all of our results remain in a version of the model that determines the size of the professional sector endogenously.

⁴Press 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.

Assume that the planner must pay a per period cost $\kappa(q)$ to create and maintain an additional position in the professional sector given that q positions already exist. We assume that $\lim_{q \rightarrow 0} \kappa(q) = 0$ and $\lim_{q \rightarrow 1} \kappa(q) = \infty$. Now, the planner makes the same decisions as before, but the planner also determines the size of the professional sector directly. The new planner's problem is

$$V(\rho^u, \rho^x) = \max_{n, \alpha^u, \alpha^x, q} v^o(2 - q) + \int_0^q \kappa(y) dy + (q - \alpha^u \rho^u - \alpha^x \rho^x) v_0^a(n) + \alpha^u \rho^u v_1^a + \alpha^x \rho^x v_1^p + \beta V\left([q - \alpha^u \rho^u - \alpha^x \rho^x] \left[1 - \frac{x}{\varepsilon} n\right], [q - \alpha^u \rho^u - \alpha^x \rho^x] \pi \frac{x}{\varepsilon} n\right)$$

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. Hence, $V(\rho^u, \rho^x)$ is still linear, and the results presented above still hold. The first order condition for q pins down the size of the professional sector, but the size of the professional sector does not impact optimal work loads for new associates or the mapping between output signals and job assignments among experienced associates.

4.2. Decentralization. There are no information asymmetries in our problem. Workers have no private information about their abilities or their actions. All output signals and all actions are public. Thus, it is not surprising that many different market mechanisms could implement the solution to our planner's problem. Appendix B proves that one particular mechanism implements our planner's solution. Here, we discuss how and why this mechanism would work.

Consider a large set of potential employers who may hire workers to produce in the outside sector or in the professional sector. Employers pay no cost to enter either sector. Employers who hire workers to produce in the outside sector pay no costs other than the wages they pay workers. Employers who hire workers to produce in the professional sector incur a cost $\kappa(q)$ per worker they hire. Recall that q is the total number of workers in the professional sector. We assume that each employer in the professional sector employs a trivial fraction of all professionals. Thus, each employer treats $\kappa(q)$ as given because he correctly believes that his own hiring decisions have negligible effects on total professional employment.

Now, assume that employers in both sectors offer contracts of the following form: Outside sector employers offer a standard contract that requires workers to perform the socially efficient work load and receive their expected output as compensation. Professional sector employers offer new workers an associate contract that requires them to perform the socially efficient work load, \hat{n} , and receive their expected output minus $\kappa(\hat{q})$ as compensation, where \hat{q} is the planner's solution for total professional sector employment. Professional sector employers offer experienced professionals a menu of two contracts. The experienced associate contract specifies a work load and compensation level as a function of the market's posterior beliefs about the ability of an experienced professional. Conditional on these beliefs, employers specify work loads that maximize expected surplus and offer compensation equal to expected output minus $\kappa(\hat{q})$. Professional sector employers also offer partner contracts that take the same form.

Because the contracts we describe dictate that workers receive the expected net output they create in a given position, all employers earn zero expected profits if they offer these contracts and the resulting choices of workers generate total professional employment equal to \hat{q} . Further, if professional sector employment is

\hat{q} , new workers should be indifferent between working in the outside sector versus the professional sector. To see this, note that the contracts we propose should induce efficient job assignment among experienced workers because these contracts transfer all expected net output to workers. Further, the first order condition for q states that, if $q = \hat{q}$ and the planner assigns experienced workers to positions optimally, the surplus new workers expect to generate over their careers is the same, whether they begin their careers in the outside sector or the professional sector.

If professional employment were less than \hat{q} , the cost of maintaining professional positions would be less than $\kappa(\hat{q})$, and new workers would strictly prefer the new associate contract. So, professional employment would increase. In contrast, if professional employment were greater than \hat{q} , new workers would strictly prefer the outside contract, and professional employment would decrease.

Appendix B demonstrates that the contracts we describe do induce efficient assignment among experienced workers. Although these results are immediate and expected in our environment, one result merits attention. Whenever the planner's solution dictates an up-or-out rule for professional workers, either z^a is so small or $k(\hat{q})$ is so large that experienced associates of uncertain ability reject the contract offered by professional employers and choose the outside sector contract instead. Thus, up-or-out environments are ones where the productivity of professionals who are not known to be low-ability but are also not worthy of promotion to partner does not cover the costs of maintaining their positions.

The costs of maintaining professional positions, $\kappa(\hat{q})$, is endogenous. Holding z^a fixed, \hat{q} and $\kappa(\hat{q})$ are increasing functions of partner productivity, z^p . Up-or-out is optimal when, z^p , z^a , and the resulting $\kappa(\hat{q})$ are such that, all associates who realize they are not going to make partner are also willing to turn down the best offer that professional firms would make to retain them. 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 are even less productive than experienced professionals. So, they are also not able to generate enough surplus in the professional sector to cover their outside option. New associates are nonetheless willing to pay a first-period utility cost since their work as new associates may reveal that they are worthy of promotion to partner.

On the other hand, once an associate learns that she is not going to be promoted to partner, employment in the outside sector dominates continued employment under the terms of her original new associate contract, and in up-or-out settings, outside employment also dominates that best experienced associate contract that professional firms are willing to offer.

The fact that working as a new associate involves paying a up-front utility cost in exchange for the possibility of winning a promotion contest may shed light on some 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.

The contracts we describe in Appendix B do not mandate dismissals. In environments where our model implies that up-or-out is optimal, experienced professionals of uncertain ability choose the outside contract over an experienced associate contract that offers to pay them their expected net output in the professional sector. 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.

4.3. Our Contribution. Several scholars have argued that up-or-out rules are puzzling because some professionals who are not well matched with the partner role are nonetheless competent workers, and their firms could offer them the opportunity to remain in their current job at a wage commensurate with their productivity. This line of reasoning makes sense if the screening process is a passive one that has simply revealed that some associates are not well-suited to the partner role. However, when the screening process involves work loads beyond static efficiency, associates who learn they are not going to be partners would never want to remain in their original jobs even if they were not asked to take a wage cut.

In our framework, the long hours that new associates work guarantee that every experienced associate who remains in the professional sector must work under new terms. When z^p is high enough relative to z^a , new associate hours are particularly long, only professionals with known high-ability are retained, and all retained professionals become partners.

In practice, adherence to up-or-out rules is not always strict. Large firms that have a small number of non-partner positions for senior specialists still hire associates who expect that they are mostly likely going to be promoted all the way to partner or asked to leave. Associates who enter firms with large numbers of senior specialists have different expectations. We argue that details of technologies and organizations that determine the productivity of these senior specialists influence the work loads that new associates bear and also shape their expectations concerning how strictly their firms will adhere to up-or-out rules.

5. EMPIRICAL PATTERNS CONCERNING HOURS AND PROMOTIONS

WORK IN PROGRESS

We have preliminary results from two different data sets that suggest the following patterns in law.

- (1) When lawyers in “big” law firms do not make partner and decide to move off the partner track into the role of Of Counsel, they bill their time at the same rate as Partners and at higher rates than associates, but they work much less than associates on the partner track. The fact that hours drop but billing rates do not when these switches occur suggest that the long hours associates work reflect more than attempts to build human capital at the beginning of careers. Once these lawyers are “off” the partner track, they are no longer “trying out” and their hours decrease
- (2) In a second data set – that follows individual lawyers – we find that hours always decrease when associates who do not make partner leave private law for other careers.
- (3) We also find in both data sets that hours change little when new associates make partner. This is possible in our model but hard to reconcile with rat-race models, where partners who have passed the screening stage should work less than the statically optimal levels.

We are also beginning to work with a third data set on MBAs that allows us to compare hours for new associate in professional partnerships with MBAs from the same school who entered corporate jobs.

6. CONCLUSION

We plan to write this when we have completed the empirical work.

7. APPENDIX MATERIAL

We have formal proofs for all our results, but they are in a long document that goes through all our analyses in mind-numbing detail. We still have to break this into pieces that would form a pithy Appendix for the paper.

REFERENCES