Allocating Effort and Talent in Professional Labor Markets

Gadi Barlevy, *Federal Reserve Bank of Chicago*

Derek Neal, *University of Chicago and National Bureau of Economic Research*

In many professional service firms, new associates work long hours while competing in up-or-out promotion contests. Our model explains why. We argue that the productivity of skilled partners in professional service firms (e.g., law, consulting, investment banking, and public accounting) is quite large relative to the productivity of their peers who are competent and experienced but not well suited to the partner role. Therefore, these firms adopt personnel policies that facilitate the identification of new partners. In our model, both heavy workloads and up-or-out rules serve this purpose.

I. Introduction

Many professional service firms employ two personnel practices that are uncommon in other labor markets. They ask young professionals to work

We thank John Kennan, Charles Lewis, Kevin Murphy, Canice Prendergast, Chris Taber, and Fabrice Tourre for useful comments. We thank seminar participants at the University of Chicago, University College London, Edinburgh University, University of Wisconsin–Madison, the 2016 Annual Meetings of the Society of Economic Dynamics, and the 2014 North American Summer Meeting of the Econometric Society as well as the National Bureau of Economic Research Organizational Economics and Personnel Economics working groups for valuable feedback. We thank Nicki Bazer, Ted Froum, Hank Kelly, and Stephanie Scharf for useful conversations about personnel practices in law firms. We thank Natascha Geirhos and Henry Doorn for helpful insights concerning the public accounting industry. We thank Andrew Jordan, Caitlin McCarthy, Jorge Luis García, Max Samels, and Ezra Karger for excel-

[*Journal of Labor Economics*, 2019, vol. 37, no. 1]

0734-306X/2019/3701-0005$10.00
Submitted March 3, 2017; Accepted October 20, 2017; Electronically published October 19, 2018

exceptionally long hours, and they require them to compete in up-or-out promotion contests. We develop a model that explains why.

Our work fills a hole in the existing literature on professional labor markets. Rat race models and the literature on career concerns provide reasons that young professionals may work long hours, but these literatures do not directly address why many professional service firms adopt up-or-out promotion rules. The literature on up-or-out explains how firms can use this policy to solve commitment problems or to facilitate the identification of talented professionals who will succeed as partners,[1] but these models almost always ignore worker effort and make no clear predictions about the effort levels of young professionals relative to those of other young workers.

While heavy workloads and up-or-out rules may serve many purposes, our results suggest that these two practices are not separate phenomena. Both practices facilitate the identification of the talented professionals who can function effectively as partners. Firms are able to learn more quickly about their new associates when they require them to perform more tasks. Furthermore, when firms replace experienced associates with new ones, they gain opportunities to identify talented professionals who will have long careers as partners.

Our model contains no hidden actions or private information. Therefore, to aid exposition we derive our key results by solving and analyzing a planning problem. We then show how to decentralize the solution to this problem. Finally, we document several patterns in data on earnings and hours worked among lawyers that support our contention that professional service firms require new associates to take on heavy workloads while participating in up-or-out promotion contests because both practices speed the discovery of new partners.

II. Literature Review

New associates in elite professional service firms (e.g., leading firms in law, consulting, public accounting, and investment banking) often work much longer hours than most white collar workers who have similar levels

lent research assistance. Contact the corresponding author, Derek Neal, at d-neal@uchicago.edu. Information concerning access to the data used in this paper is available as supplemental material online.

[1] We use the term "partner" to denote a leader who is responsible for business development and resource allocation regardless of the ownership structure of her firm. Some consulting firms, e.g., Accenture, and most investment banking firms are no longer organized as partnerships. Yet young professionals in these firms continue to take on heavy workloads, and these firms continue to follow up-or-out promotion rules. Modern investment banks are involved in many lines of business other than professional services. Our model is most applicable to groups of investment bankers who provide advice and services related to mergers and acquisitions. In these groups, up-or-out is the norm.

of education. Building on the work of Akerlof (1976), Landers, Rebitzer, and Taylor (1996) offer a possible explanation for this pattern. They treat law firms as teams and assume that while team output is observed, individual output is not. Partners benefit from hiring associates with low effort costs because this allows them to sell their equity shares to more productive lawyers in the future. In the separating equilibrium that Landers, Rebitzer, and Taylor (1996) describe, a menu of employment contracts specifies hours requirements and compensation for new associates in each law firm. Lawyers who share the same effort costs select the same contracts and work together in the same firms. Ex post, associates in all firms work more than the efficient number of hours.[2]

In Holmstrom (1999), employers do not observe the actions workers take, and output-contingent contracts are not possible. In this setting, Holmstrom (1999) shows that young workers may choose more than the efficient level of effort in order to shape market beliefs about their abilities. In equilibrium, firms infer 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.

The up-or-out literature contains several variations on two different approaches, but neither approach addresses workloads for associates. One literature characterizes up-or-out rules as commitment devices that solve a double moral hazard problem between workers and firms. In these papers, 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 renege on payments linked to performance signals that are not verifiable in court. Given up-or-out promotion rules, firms lose valuable workers whenever they fail to promote those who produce positive output signals, and these potential losses make promises to reward hidden actions more credible.

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.[3] Prendergast (1993), Waldman (1990), and Ghosh and Waldman (2010) extend the logic of Kahn and Huberman (1988) to settings where various features of the employment relationship point to a specific reason that up-or-out rules may or may not make contingent promises concerning raises and promotions more credible.[4]

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

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

[4] Ghosh and Waldman (2010) also model worker effort. However, in contrast to our model, worker actions are hidden in their model. Also, in their model effort

In Rebitzer and Taylor (2007), up-or-out rules help enforce a commitment to fix an overall ratio of associates to partners. This maintains partnership stability by raising the cost that any group of partners would face if they decided to take their clients and form a new firm.

Both O'Flaherty and Siow (1992) and Demougin and Siow (1994) link up-or-out rules to optimal screening procedures. Demougin and Siow (1994) do so in a model of hierarchies where 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.

Our results link up-or-out rules to the option value logic found in O'Flaherty and Siow (1992) and Demougin and Siow (1994), but our results also differ in two ways. First, we link the number of tasks that young professionals perform to the information that the market receives about them. Second, our comparative static results concerning when up-or-out regimes exist do not deal with changes in outside options. Rather, we focus on changes in the relative productivities of experienced professionals who occupy different roles.

Some scholars argue that up-or-out rules are puzzling, since such rules likely mandate the dismissal of workers who are performing well in their current positions, but we argue that there may be no puzzle to explain. The long hours that associates work are a cost that young professionals pay to learn whether they are well suited to become partners. If these young professionals learn that they are not going to become partners, they are no longer willing to pay this cost—that is, they would not continue doing their current job even if their firms allowed them to do so. Furthermore, since they are not suited to the partner role, their best options for future employment typically lie elsewhere.

Finally, our model provides new insights concerning life-cycle patterns of changes in hours worked among professionals who follow different career trajectories. We document that among lawyers with roughly 10 years of experience in private law firms, those who leave private law or leave the partnership track within private law reduce their hours significantly even though their wage rates often rise. This pattern is difficult to understand in most models of life-cycle labor supply. However, in our model the heavy workloads that young law associates take on are a key component of their audition for partnership positions. When an associate learns that she is not going

levels among new professionals may be high or low in both up-or-out firms and firms that follow standard promotion practices.

to make partner, her audition is over, and she may well reduce her hours even though her skill set continues to command a high wage rate.

III. Model Setup

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

A. 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 2 periods. Thus, in any period a mass 2 of workers exists.

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

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

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

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

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

where m is expected income, n is the number of tasks performed, and $c(n)$ is the disutility of performing n tasks. We assume $c(0) = 0, c'(0) = 0$. Furthermore, there exists a maximum workload, $\bar{n}$, such that $\lim_{n \to \bar{n}} c'(n) = \infty$, $c''(n) > 0 \ \ \forall\, n \in [0, \bar{n}]$. All workers pay the same utility cost to complete any task.

Let θ denote worker ability, which is either high or low—that is, $\theta \in \{0, x\}$, 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.

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

$$y^o = w^o n + \epsilon. \tag{1}$$

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

$$y_0^a = (1 + \theta)n + \epsilon. \tag{2}$$

The production function for experienced associates is

$$y_1^a = z^a(1 + \theta)n + \epsilon. \tag{3}$$

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

$$y_1^p = \begin{cases} z^p(1 + \theta)n + \epsilon & \text{if } \theta = x, \\ -\infty & \text{if } \theta = 0. \end{cases} \tag{4}$$

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

Our planner must allocate workers between the professional labor market and the outside sector. For now, we cap employment in the professional sector at $q < 1$ to capture the idea that only a fraction of highly educated agents work in the professional sector. Later, we treat q as an endogenous variable that is determined by the costs of maintaining professional jobs and the productivities of positions in the professional sector.

We impose the following restrictions on production parameters:

$$z^a < w^o < z^a(1 + \pi x). \tag{5}$$

The first inequality in expression (5) implies that an experienced associate who has low ability is more productive in the outside sector than in the professional sector. The second inequality implies that the expected productivity of an experienced associate with unknown ability is greater in the professional sector than in the outside sector. If this were not true, the planner's assignment decisions for experienced professionals with unknown ability would be trivial.

B. Learning

The market observes the output that each worker produces in each period. In Barlevy and Neal (2016), we show that given a specific assumption about the distribution of ϵ, the output signal for any new associate either fully reveals her talent or provides no information about her talent. Furthermore, the likelihood that an output signal is fully revealing increases with n, the number of tasks an associate performs.

The specific learning technology in Barlevy and Neal (2016) is not essential for our analysis, but we do require all informative signals to be fully revealing, that is, the market either infers an associate's type perfectly or learns nothing about the associate. Define $\phi(n)$ as the probability that an associate who performs n tasks produces a signal that fully reveals her type. We assume that $\phi(n)$ is an increasing and concave function of n, that is, $\phi'(n) > 0$, $\phi''(n) \leq 0$, and $\phi(\bar{n}) < 1$.[5]

IV. The Planner's Problem

The planner seeks to maximize the expected present discounted value of the sum of present and future differences between per-period output and effort costs by assigning workers of different types to jobs and workloads. In this economy, the nature of a worker's past job experience and the market's information about a worker's talent define a worker's type.

Five different types of workers exist. Experienced workers who spent their first period working as associates may be known to be low ability or high ability, or their first-period output may not have revealed their type. Because work in the outside sector produces no signals about worker ability, the market has a common belief about the ability of all workers who possess

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

1 period of experience in the outside sector. Finally, the market has the same common belief about the ability of all workers who have no experience.

In each period, the planner must assign a position and a workload to each of these five worker types. Appendix A demonstrates that all but two of these choices are immediate given our assumptions; for example, with the possible exception of the initial period, the planner always dismisses experienced professionals with known low ability and always promotes those with known high ability. The two nontrivial choices are how many tasks, n, to give to new associates and what fraction, α^u, of experienced associates with unknown ability to retain. If the planner retains none of these associates—that is, $\hat{\alpha}^u = 0$—we say that the planner follows an up-or-out rule.

The state space for the planner's problem can be reduced to two variables: the stock of experienced professionals with unknown ability, ρ^u, and the stock of experienced professionals with known high ability, ρ^x.

We use the following notation for the surplus generated by assigning different types of workers to different jobs:

- v^o is the surplus generated by an outside worker; and
- v^j_s is the surplus generated by a professional worker with $s \in \{0, 1\}$ periods of experience in position $j \in \{a, p\}$.

Appendix A establishes that the per-period surplus generated by the planner's current choices equals

$$s(n, \alpha^u, \rho^u, \rho^x) = (2 - q)v^o + (q - \alpha^u\rho^u - \rho^x)v^a_0(n) + \alpha^u\rho^u v^a_1 + \rho^x v^p_1. \quad (6)$$

The first term represents surplus generated in the outside sector. The second term captures surplus generated by new associates. The last two terms capture surplus generated by experienced professionals. We write $v^a_0(n)$ to stress that workloads for new associates are one of the two key choices for the planner. For all other workers, optimal workloads are solutions to simple static maximization problems.

Note that there can be at most q new associates in any period, and the planner always learns that a fraction of these new associates have low ability. Thus, in any period after the initial period, $\rho^u + \rho^x < q$. Here, we discuss the planner's optimal choices in this region of the state space.

If we assume that the planner discounts the future using $\beta < 1$, we can write the planner's problem using the following recursive formulation:

$$\begin{aligned} V(\rho^u, \rho^x) &= \max_{\alpha^u, n} s(n, \alpha^u, \rho^u, \rho^x) + \beta V(\rho^u_+, \rho^x_+), \\ \text{s.t.} \qquad \rho^u_+ &= [q - \alpha^u\rho^u - \rho^x][1 - \phi(n)], \\ \rho^x_+ &= [q - \alpha^u\rho^u - \rho^x]\pi\phi(n). \end{aligned} \quad (7)$$

If the stock of experienced professionals with unknown ability is positive—that is, $\rho^u > 0$—the planner faces a trade-off. These workers are more productive than new associates because they have more professional experience. However, they are not qualified to be partners, and for each one the planner retains today he will have one less experienced professional next period, which means he will have fewer partners next period. Furthermore, the planner's effort choice for new associates, n, interacts with these retention decisions because the probability that the planner observes the actual ability of a new associate is $\phi(n)$.

Recall that there are no productivity spillovers among workers in this model. We therefore expect $V(\rho^u, \rho^x)$ to be linear. Appendix A shows that this is true when $0 \leq \rho^u + \rho^x < q$.

Claim 1: $V(\rho^u, \rho^x) = K_1 + K_2\rho^u + K_3\rho^x \;\; \forall \, (\rho^u, \rho^x)$ such that $0 \leq \rho^u + \rho^x < q$.

Substitute this expression for $V(\rho^u, \rho^x)$ into equation (7) and take the derivative with respect to n. This yields the first-order condition that defines the optimal workload for new associates, $\hat{n}$:

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

The marginal cost of new associate effort must equal the sum of two marginal returns. The first, $(1 + \pi x)$, is the expected marginal product of new associate effort. The second, $\beta\phi'(n)(\pi K_3 - K_2)$, is the marginal information return from worker effort. For the planner, K_2 represents the value created by replacing a new associate with an experienced professional who possesses uncertain ability, while K_3 represents the value created by replacing a new associate with an experienced professional who has known high ability. When the planner marginally increases a new associate's workload, n, the probability that the associate's output signal reveals her ability increases by $\phi'(n)$. If the signal is revealing, there is a probability π that the signal will reveal high ability, and the planner's value function will increase by K_3 instead of K_2. Thus, $\phi'(n)(\pi K_3 - K_2)$ is the marginal information rent generated next period by new associate work this period.

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

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

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

A similar trade-off between current output and future information shapes the planner's decision concerning the retention of experienced associates of uncertain ability. In appendix A, we solve for K_2 and K_3 and use these solutions to express the first-order condition for α^u as follows:

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

The left-hand side of the inequality in proposition 2 is the current surplus cost of replacing an experienced associate of unknown ability with a new associate. The right-hand side gives the expected future returns that these replacements create. The probability that new associates become partners next period is $\pi\phi(\hat{n})$, and the additional surplus generated by partners is $v_1^p - v_1^a$. Thus, the planner's decision concerning $\hat{\alpha}^u$ reflects a trade-off between the current value of worker experience in the associate position and the discounted expected value of identifying more high-ability partners. Up-or-out rules, $\hat{\alpha}^u = 0$, are optimal in environments where the increase in expected future surplus associated with increasing the number of future partners outweighs the loss in current surplus that comes from replacing experienced associates with new ones.

Because workers live for only 2 periods in our model, all promotions and sector changes take place at the end of period 1. However, the fact that all promotions and dismissals take place at a common experience level is not an essential feature of the up-or-out equilibrium we characterize. Informative signals are fully revealing in our model. Therefore, if workers in our model lived for $S > 2$ periods, the planner would still assign workers to the outside sector whenever he learned that they possessed low ability, and the planner would still promote workers to the partner position whenever he learned that they possessed high ability. Moreover, as long as S is finite, the option value of associates with unknown ability declines as they gain more experience. This implies that the planner may choose to replace all associates who work $\hat{s} < S$ periods without producing a signal that reveals their true ability. For us, the existence of this time limit, $\hat{s} < S$, is a defining feature of an up-or-out regime.[6]

O'Flaherty and Siow (1992) define up-or-out differently. In their model, there is no fixed date $\hat{s}$ that constrains the timing of promotions and dismissals. Because workers live forever, the option value of an associate does

[6] In our work with the After the JD survey (see app. D), we have learned that among law associates, departures and promotions occur at many different tenures. However, only a trivial number of new associates work as associates for more than 12 years in the same firm.

not decline with experience. Still, because experience produces information, all workers eventually move up to the partner role or leave their firm.[7]

A. Comparative Statics

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

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

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

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

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

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

Taken together, propositions 3 and 4 imply that in a region of the parameter space, both up-or-out promotion rules and heavy workloads for new associates are optimal. Given any initial pair (z^a, z^p), $\hat{n}$ increases and an up-or-out rule is either retained or adopted as we increase z^p. Furthermore, $\hat{n}$ decreases and an up-or-out rule is either retained or abandoned as we increase z^a.

[7] The signals in O'Flaherty and Siow (1992) are not fully revealing, so standards for promotion depend not only on beliefs about the value of associate talent but also on the precision of these beliefs.

Several authors point out that strict adherence to up-or-out rules became less common in law firms during the 1980s and 1990s.[8] Gorman (1999) argues that mergers, changes in information technology, and increases in the demand for specialized legal services allowed larger law firms to create a limited number of new nonpartner roles for experienced specialists. Among lawyers who were well suited to these new roles, z^a increased, and it became optimal to retain these lawyers even if they were not well suited to the partner role.

Growing relative demand for specialists has produced similar movements away from strict up-or-out policies in public accounting, especially in the largest accounting firms.[9] This development in public accounting is of particular interest because it parallels a trend toward less demanding work schedules in public accounting. In recent decades, large public accounting firms have adopted a number of practices that seek to reduce work-related stress. Many firms have increased paid holidays, the use of job-sharing arrangements, and the use of flexible scheduling.[10]

It is difficult to investigate the extent to which our model provides the correct explanation for the recent creation of more permanent nonpartner roles for experienced professionals in law and accounting or for the contemporaneous adoption of policies that promote work-life balance. Propositions 3 and 4 above deal with changes in z^a or z^p holding all else constant, but the waves of mergers and changes in information technology that affected law and accounting firms in recent decades likely affected the productivity of new associates, experienced specialists, and partners simultaneously.

The predictions of our model map more cleanly into long-standing differences between the labor market for professional services and other markets for well-educated workers. The trade literatures on professional labor markets clearly indicate that most new entrants in professional service firms continue to compete in up-or-out promotion contests while working more arduous schedules than their peers in other industries.[11] Our model suggests that this is because an entry-level professional in a major manufacturing or service firm is not really auditioning to be chief executive officer or even a division president. While the largest professional service firms have thousands of partners, the largest traditional corporations have only handfuls of people in their highest leadership roles. Furthermore, while professional service firms offer a limited number of senior roles for nonpartners, large cor-

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

[9] See the *New York Times*, May 17, 1990. See Almer (2004) for more recent data from the American Institute of Certified Public Accountants.

[10] See table 23 in Almer (2004). See also Greenhouse (2011) and Lewison (2006).

[11] In addition, Bertrand, Goldin, and Katz (2010) report that recent MBA graduates who enter investment banking and consulting appear to work longer hours than their peers who enter corporate jobs.

porations contain many productive roles for experienced professionals who do not reach the absolute highest levels of management.

Viewed through the lens of our model, an experienced professional in a traditional firm enjoys a large effective value of z^a because the firm can assign her to the one role, among many potential roles, that suits her best. In contrast, professional service firms are horizontal organizations that grow by adding partners who can attract and maintain clients. A new entrant in a professional service firm is auditioning to be such a partner and expects to leave the firm if the audition for that one role does not go well. Thus, it makes sense to expedite the process of determining whether this new entrant is actually partner material.[12]

Young workers acquire not only information but also new skills through work experience. In appendix B, we present a related model where associates acquire skills by performing tasks, but they do not know ex ante how fast they will learn. In this model, new associates take on heavy workloads in an attempt to reach the skill level required to work as a partner, and if the productivity of partners is great enough relative to the productivity of those at medium skill levels, the planner replaces all experienced associates who do not reach the high skill level with new associates who may become future partners. Whether new associates are learning about their skill level or learning about their capacity to acquire skills, our key insights remain.

V. Endogenous Sector Size and Decentralization

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

A. Endogenous Sector Size

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

Professional workers require office space, support staff, and other resources that facilitate their capacity to interact efficiently with clients. We assume that

[12] A large literature points out that, in models with one position, the option value associated with a new worker determines the stringency of the rule that governs retention decisions. Our assumptions concerning z^p, z^a, and the losses that would occur if low-ability types were to occupy the partner position highlight the fact that our model is not a one-position model. We model an up-or-out decision, not a stay-or-go decision.

the supply curve for effective support services to the professional sector is upward sloping. As the professional sector grows, it must employ support staff who have better outside options and occupy office space that has more valuable alternative uses.

Suppose the per-period cost of supporting the qth professional position is given by $\kappa(q)$, where $\kappa(\cdot)$ is an increasing continuous function such that $\lim_{q\to 0}\kappa(q) = 0$ and $\lim_{q\to 1}\kappa(q) = \infty$.[13] Given our previous assumptions, these restrictions ensure that the planner assigns some but not all of the workers in a cohort to begin their careers in the professional sector.

The planner faces a problem similar to the one described in equation (7), but here he must also determine $\hat{q}$, the optimal number of professional workers. Let α^x denote the fraction of experienced professionals with known high ability that the planner retains. Appendix C shows that, with the possible exception of the initial period, the planner always retains these workers and promotes them to partner, just as in the fixed-q problem.[14] However, explicit notation for this decision allows us to more clearly formulate our new planner's problem:

$$V(\rho^u, \rho^x) = \max_{n,\alpha^u,\alpha^x,q} s(n, \alpha^u, \alpha^x, q, \rho^u, \rho^x) - \int_0^q \kappa(y)\,dy + \beta V(\rho^u_+, \rho^x_+), \tag{8}$$

$$\text{s.t.} \quad \begin{aligned} &(1) \quad \rho^u_+ = [q - \alpha^u\rho^u - \alpha^x\rho^x][1 - \phi(n)], \\ &(2) \quad \rho^x_+ = [q - \alpha^u\rho^u - \alpha^x\rho^x]\pi\phi(n), \\ &(3) \quad q - \alpha^u\rho^u - \alpha^x\rho^x \geq 0. \end{aligned} \tag{9}$$

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

Appendix C shows that, with the possible exception of the initial period, the planner chooses an optimal sector size, $\hat{q}$, that is constant over time. Appendix C also shows that if the initial stocks of experienced professionals are so large that the planner chooses a sector size greater than or equal to $\hat{q}$ in

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

[14] If the initial stock of skilled professionals, ρ^x_0, is sufficiently large, the planner may not retain all of them in the professional sector. However, we show in app. C that, after the initial period, the planner promotes all skilled professionals to partner.

the initial period, constraint (3) will bind, and the planner will assign all new workers to the outside sector in the initial period. Taken together, these results imply that there can be at most $\hat{q}$ new associates in any period. Hence, beyond the initial period the stock of experienced workers can be at most $\hat{q}$, and some will be known to possess low ability. This implies $\rho^u + \rho^x < \hat{q}$, so constraint (3) never binds after the initial period.

Since the productivity of a worker in a given position is not influenced by how other workers are allocated to positions, optimal workloads, retention rules, and promotion policies are the same whether there are five professional positions or five million. For this reason, the planner chooses $(\hat{n}, \hat{\alpha}^u, \hat{\alpha}^x)$ using the same rules he employed in the fixed-q problem, and all of the comparative statics results that we derived in Section IV continue to hold.

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

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

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

B. Decentralization

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

Assume that all workers choose whether to work in the outside sector at a fixed wage, w^o, or work in the professional sector as independent contractors.[15] Whether these independent professionals choose to work in the as-

[15] We treat workers as independent contractors to facilitate exposition. The same results would hold in a competitive labor market where identical professional service firms posted a menu of employment contracts that specified optimal workloads, ter-

sociate or partner role, we assume that each must hire support services at a cost determined by the equilibrium size of the professional sector.[16]

Appendix C demonstrates that the market equilibrium in our independent contractor scenario implements the planner's solution to the problem described by equation (8). Here, we sketch the argument in two steps. First, we consider how independent contractors choose workloads and jobs when they believe the sector size is going to equal the planner's choice $\hat{q}$. Then we argue that if all contractors follow these rules, rational entry decisions will generate a professional sector of size $\hat{q}$.

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

$$\begin{aligned} v(n, q) = {} & v_0^a(n) - \kappa(q) + \beta\pi\phi(n)[v_1^p - \kappa(q)] \\ & + \beta(1 - \pi)\phi(n)v^o + \beta(1 - \phi(n))\max[v^o, v_1^a - \kappa(q)]. \end{aligned} \tag{10}$$

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

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

Next, consider the last term of $v(n, q)$ in equation (10). Appendix C also demonstrates that the planner imposes an up-or-out rule, $\hat{\alpha}^u = 0$, if and only if experienced independent professionals with uncertain ability choose the outside sector when facing a professional sector size of $\hat{q}$, that is, $v^o > v_1^a - \kappa(\hat{q})$. To gain intuition for this result, recall that, without regard to sector size, the planner chooses $\hat{\alpha}^u = 0$ in all environments where new associates are more valuable than experienced associates. Furthermore, the plan-

mination rules, and wages equal to expected marginal products for all possible combinations of worker types and positions.

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

ner's choice of $\hat{q}$ and therefore $\kappa(\hat{q})$ equates the value of a new associate and an outside worker. If outside workers are just as valuable as new associates and new associates are more valuable than experienced associates, we expect an experienced professional who is not partner material to find outside work more attractive than the experienced associate position.

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

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

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

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

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

Our model treats new associate work as an audition for the partner role, and in our contractor economy the price of support staff clears this market for auditions. However, $\kappa(\hat{q})$ is not the only audition cost that new associates pay. New associates also choose to take on heavy workloads, and as a result they enjoy lower first-period utility than outside workers.[17]

This insight may help explain why young professionals often report low job satisfaction and, in particular, why they report that they would be willing to accept lower earnings in exchange for less demanding work schedules.[18] Suppose that young professionals who respond to such surveys are reporting that they are willing to accept lower earnings in exchange for less demanding workloads, holding all else constant, including their future prospects for promotion. If so, new associates in our model would express a willingness to exchange current salary for reduced workloads. The problem is that there is no way to make such an exchange while holding all else constant. If new associates did perform fewer tasks, the market would learn less about them, and they would be less likely to become partners.

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

Some scholars have argued that up-or-out policies are puzzling because surely some experienced associates who are not well suited to the partner role are nonetheless competent professionals. Why would firms refuse to negotiate a retention package for these associates?[19] Our interpretation of associate jobs as auditions for partnerships provides an answer. In our framework, young professionals in up-or-out markets pay a utility cost to acquire information about whether they are well suited to lucrative partner positions. Thus, when a young professional learns that she is not partner material, she is not willing to pay this cost any longer. Furthermore, although she would be more productive, in expectation, than the new associate who replaces her, her firm

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

[18] See Landers, Rebitzer, and Taylor (1996) for survey responses from young lawyers.

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

cannot profitably make a retention offer that she would accept because it can no longer offer her the opportunity to learn about her fitness for the partner role.

VI. Empirical Patterns concerning Hours and Promotions

Next, we explore data that describe career outcomes among young lawyers. Before we describe our results, we must discuss several issues that arise concerning the mapping between the positions in our model and the job titles we observe in our data. In our model, there are only two positions, associate and partner. For much of the twentieth century most private law firms created only these two positions, but in recent decades many firms have created additional positions. Many firms now have nonequity partner positions, and a smaller number have "of counsel" or counsel positions.

We proceed under the assumption that persons who are promoted from associate to nonequity partner early in their careers are likely persons who are still trying to earn promotions to full partner, but we assume that persons who transition from associate to counsel positions early in their careers are no longer under consideration for promotion to partner.[20] Nonequity partner appears to be a stepping stone to partner, but in many firms where up-or-out is the rule, counsel positions appear to exist as exceptions to the rule. According to Flom and Scharf (2011), most law firms report that their counsel attorneys are not eligible for promotion to partner. Furthermore, data that we describe more fully in appendix D suggest that young attorneys who accept counsel positions rarely return to partner-track positions. Both the academic literature and the trade literature suggest that persons who occupy counsel positions often do not possess the business development skills that partners need but do possess special expertise that allows them to be exceptionally productive in a nonpartner role.[21]

Table 1 describes data from the Survey of Law Firm Economics (SLFE), which is conducted annually by ALM Legal Intelligence. The data come from eight annual surveys taken during the period 2007–14. Some firms appear in more than one annual survey, but these are not observations from a panel data set. Rather, the data come from eight repeated cross-sectional surveys.[22]

Table 1 describes outcomes for lawyers who are between 8 and 12 years into their careers. We chose this experience interval because law firms now

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

[21] In the terms of our model, these are persons with an effective z^a that is greater than the standard level.

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

Table 1
Hours Billing, Billing Rates, and Total Compensation By Position

	Partner	Nonequity Partner	Associate	Counsel
Average hours billed	1,646	1,612	1,493	1,128
	(416)	(538)	(592)	(633)
N	2,982	3,135	7,144	558
Average hourly rate	290	299	259	304
	(77)	(97)	(85)	(105)
N	2,990	3,131	6,931	543
Average compensation	233,970	197,471	143,409	144,305
	(110,968)	(78,307)	(53,157)	(67,458)
N	3,053	3,188	7,283	574

NOTE.—This table reports job characteristics for lawyers included in the Survey of Law Firm Economics conducted by ALM Legal Intelligence between the years 2007 and 2014. Shown for each characteristic are the sample mean, the standard deviation (in parentheses), and the sample size (*N*). This table considers only attorneys with between 8 and 12 years of experience, defined as years since passing the bar. Compensation is defined as take-home salary and retirement contributions plus year-end bonus plus benefits. Hours billed is the annual number of billable hours for each attorney. Hourly rate is the typical hourly rate charged by each attorney. Sample sizes differ among cells because of different frequencies of item nonresponse.

make many crucial retention and promotion decisions in this interval. The table presents results for associates, equity partners, nonequity partners, and counsel attorneys. Because these counsel attorneys are less than 12 years into their careers, it seems reasonable to assume that most of them are attorneys who recently left an associate position in their current or previous firms and now occupy a senior role off the partnership track. Thus, it is interesting to note that, compared with the other three groups, counsel attorneys bill fewer hours but charge clients higher rates for their time. For example, counsels bill their time at rates 17% greater than associates yet bill 24% fewer hours.

Why would associates bill so many more hours than their peers who work as counsels? Our model suggests that this hours gap reflects the fact that counsel attorneys are no longer auditioning for partnerships. Their past work experience has revealed their type, and they are no longer producing signals about their suitability for the partner position.[23]

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

The SLFE is a cross-sectional survey, so we cannot know with certainty that the counsel attorneys in our SLFE sample recently left associate positions. Tables 2 and 3 present results from a second data set that provides much smaller samples but does allow us to track individual lawyers over time.

[23] Gender is missing for many records in the SLFE data. However, in the sample of respondents who report being male, the same patterns exist.

Table 2
After the JD (AJD) Survey Full Sample: Hours Worked and Compensation by Wave 3 Position

	Partner	Nonequity Partner	Associate	Counsel	Other
Average hours wave 1	51.7	51.8	49.5	49.7	49.9
	(12.7)	(9.4)	(12.4)	(11.2)	(13.1)
N	161	103	75	31	343
Average salary wave 1	116,243	122,125	92,157	141,578	131,753
	(48,985)	(48,303)	(39,828)	(47,452)	(70,311)
N	165	107	74	32	360
Average hours wave 3	51.2	54.3	51.9	46.2	47.0
	(12.8)	(12.2)	(10.7)	(12.2)	(12.8)
N	171	109	77	32	365
Average compensation wave 3	240,154	225,653	134,187	215,975	190,566
	(220,199)	(110,588)	(75,675)	(136,151)	(197,826)
N	121	90	72	26	316

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

The After the JD (AJD) study conducted three rounds of interviews with a cohort of lawyers who passed the bar around 2000. Wave 1 interviews took place between May 2002 and March 2003. Wave 2 interviews began in May 2007 and ran through early 2008. Wave 3 interviews took place be-

Table 3
Regressions of Changes in Hours (Wave 3 − Wave 1) on an Indicator for Leaving the Partnership Track

				Firm Size ≥150	
	Full	Males	Females	Males	Females
Off-track wave 3	−3.70	−3.25	−4.26	−7.43	−5.01
	(1.15)	(1.53)	(1.78)	(2.76)	(2.92)
N	708	425	279	178	114

NOTE.—See the note to table 2 for a description of the After the JD sample. Once again, we restrict our samples to persons who report in wave 1 that they are associates in private law firms, and we eliminate five respondents who report working in staff or contract positions in wave 3. Those who are equity partners, nonequity partners, or associates in a private law firm are on-track. Those who are solo practitioners, counsels in a private law firm, or employees of an organization that is not a private law firm are off-track. The entries here are regression coefficients on a dummy variable indicating that, in wave 3, a lawyer is off-track. All regressions are bivariate regressions of changes in hours worked per week (i.e., wave 3 hours minus wave 1 hours) on the off-track indicator. Standard errors are in parentheses. *N* denotes sample size. $Y = \Delta$hours (wave 3 − wave 1). $X = 1$ if off-track in wave 3 (counsel, other).

tween May 2012 and December 2012. We chose the subsample of these lawyers who participated in both the wave 1 and the wave 3 interviews and who reported in wave 1 that they worked as an associate in a private law firm. By wave 3, most of these lawyers should have had 11 or 12 years of experience as lawyers. Thus, they are slightly more experienced than the average lawyer in our table 1 sample.

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

All of the lawyers in this sample were working as associates in private law firms in wave 1. The first row of table 2 shows that young associates who are going to make partner or nonequity partner appear to work slightly longer hours in wave 1. Since the wave 1 interviews took place in year 2 or 3 of these young lawyers' careers, these small differences in wave 1 hours may indicate that partners give more work to second- and third-year associates who, based on the quality of their early work, appear more likely to make partner.[24]

In wave 3, those who took counsel positions and the much larger sample who are no longer in private law firms work about 3 hours less per week than they worked as new associates. In contrast, those who are still trying to become equity partners work more than they worked in wave 1. Those who are partners in wave 3 work 30 minutes less per week than they worked in wave 1.

The wave 3 contrast between associates and lawyers in our other category is striking. Between waves 1 and 3 annual earnings grew for both groups by about 45%, but in wave 3 those who remain in associate positions work roughly 5 hours per week more than those who have left private law.

Table 3 reports the results from a series of bivariate regressions of hours worked in wave 3 minus hours worked in wave 1 on a dummy variable indicating whether a lawyer is off the partnership track in wave 3. Lawyers who leave the partnership track reduce their hours by roughly 3–7 hours per week relative to lawyers who remain on the partnership track. There is no evidence that this result is driven by women seeking to reduce their hours in order to spend more time at home.

The empirical results in tables 1–3 are only suggestive. The AJD surveys have response rates far below 1, and the SLFE documentation does not re-

[24] Wilkins and Gulati (1998) discuss differences in work assignments among associates. Gicheva (2013) analyzes the link between hours choices and future promotion prospects.

port response rates. Nonetheless, our results do suggest that many lawyers who abandon the partnership track in private law firms reduce their work hours even though their wage rates are constant or rising. Our model can easily generate this pattern. Furthermore, our claim that associates are auditioning for partnerships that leverage the skills of particularly gifted people squares well with two other features of the AJD data. Promotion rates are low in these markets,[25] and lawyers do not cut back significantly on their hours after they become partners.

VII. Conclusion

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

Some large law and public accounting firms no longer adhere strictly to up-or-out rules, and some, particularly in public accounting, have adopted policies that promote work-life balance. Future research should examine potential connections between these developments and the growth of large law and accounting firms. These large firms may have found ways to create new and highly productive roles for skilled experts who are not well suited to the partner role.

Appendix A

Planner's Problem and Proofs

In this appendix, we describe the planner's problem in detail and characterize its solution.

We first argue that many of the planner's choices are trivial. Table A1 lists the five different types of workers the planner faces. The planner must choose a job assignment and an effort level for each of these types. We first argue 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 solve standard static optimization problems and can also be easily characterized. We then rigorously analyze the two assignment decisions and one effort choice that remain.

[25] The vast majority of wave 1 associates in the AJD leave their wave 1 firm before wave 3, and less than 1 in 10 make partner in their wave 1 firm by wave 3.

Table A1
Optimal Assignment for Worker Types

	Ability		
History	$\theta = 0$	$\theta = x$	$\Pr(\theta = x) = \pi$
New	NA	NA	Associate/outside
Experienced outside	NA	NA	Outside
Experienced professional	Outside	Partner	?

NOTE.—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. NA = not applicable.

The rows of table A1 correspond to the three types of work experience possible for a worker. Since workers live 2 periods and can work only in one sector per period, a worker may have no experience in either sector, 1 period of experience in the outside sector, or 1 period of experience in the professional sector. The columns of table 1 correspond to the three types of information that may be available for a worker. A worker may be known to have low ability ($\theta = 0$), high ability ($\theta = x$), or uncertain ability (where $\theta = x$ with probability π and $\theta = 0$ with probability $1 - \pi$).

The combination of the three experience types and three information sets yield the nine cells in table 1. The first two columns of the first two rows are marked as not applicable since by assumption the ability of workers with no professional experience is uncertain. Only the remaining five cells correspond to types the planner actually faces.

Of these five types, three can be assigned trivially. First, new workers with no prior experience have uncertain ability and will never be asked to work as partners given $y_0^p = -\infty$. Given that production technologies are linear, the planner either assigns all of them to the outside sector or assigns some fraction of them to work as associates in the professional sector up to the point where all q positions in the professional sector are staffed.

Second, the planner should assign workers with experience in the outside sector to remain in the outside sector. Clearly, the planner would never assign them to work as partners, given the risk they might be low types. But the planner should also not assign them to work as associates, since they are dominated by inexperienced new workers; the latter have the same expected productivity working as associates, but they also have the prospect of being promoted to partner next period. Since professional employment is capped at $q < 1$, there will always be new workers the planner could draw on to replace experienced workers from the outside sector.

Third, the planner should assign workers revealed to be low ability to the outside sector. These workers are more productive in the outside sector given $z^a < w^0$, and there is nothing to gain from leaving them in the professional sector.

This leaves the two assignment decisions for experienced professionals whose ability is either uncertain or known to be high. Below, we show that after at most 1 period, the planner would assign all experienced workers known to be high ability to the partner job. This leaves the question of where to assign experienced workers with uncertain ability: should the planner leave them in the professional sector or assign them to the outside sector? We will refer to the latter case as up-or-out, since if that is the optimal assignment, all new professional workers expect that after enough time has transpired they will either be promoted up to partner or sent to the outside sector.

Finally, the planner must assign workloads to each of the five worker types. Experienced workers will be asked to put in the effort level that maximizes the surplus they create in the last period of life. New workers assigned to the outside sector will also be asked to put in that level of effort, since they will remain in the outside sector and there is no gain from distorting their effort. The only nontrivial decision is the workload of new workers employed as associates in the professional sector. For these workers, heavier workloads generate more output and more information about ability. Because this information guides assignment decisions in the next period, the optimal workload for new associates will reflect the fact that effort today affects assignments and workloads in the future.

Given these results, we can state the planner's objective function. We assume that z^a and z^p are large enough to ensure that the planner would want to fill all q positions in the professional sector. Thus, he will employ $2 - q$ workers in the outside sector, each of whom will produce a surplus of

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

In the professional sector, there are ρ^x experienced workers known to be high ability, and the planner chooses the fraction α^x of such workers to retain. The planner can use them to produce a surplus of

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

There are also ρ^u experienced workers with unknown ability, and the planner chooses the fraction α^u of such workers to retain. The planner can use them to produce a surplus of

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

Finally, the remaining $q - \alpha^u\rho^u - \alpha^x\rho^x$ workers are inexperienced workers who will be assigned to work as associates. The surplus they produce when young is given by

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

where n is the effort they put in. Hence, the flow surplus the planner creates given his choices of α^u, α^x, and n is given by

$$s(\alpha^u, \alpha^x, n) \equiv (2 - q)v^o + qv_0^a(n) + \alpha^x \rho^x (v_1^p - v_0^a(n)) + \alpha^u \rho^u (v_1^a - v_0^a(n)),$$

and the value function $V(\rho^u, \rho^x)$ satisfies the Bellman equation

$$V(\rho^u, \rho^x) = \max_{\alpha^x, \alpha^u, n} s(\alpha^u, \alpha^x, n) + \beta V(\rho^u_{+1}, \rho^x_{+1}), \tag{A1}$$

where

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

as well as the constraint

$$q - \rho^u \alpha^u - \rho^x \alpha^x \geq 0. \tag{A2}$$

This constraint states that the planner cannot retain more experienced professional workers than there are slots for them in the professional sector. Note that in the text, we express the surplus $s(\cdot)$ as a function of the state variables and substitute in $\alpha^x = 1$. We prove that, at least in a certain region, $\alpha^x = 1$ will indeed be optimal.

We now proceed to characterize the optimal path for the planner. We begin with some preliminary results and then prove the key propositions stated in the text.

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

Proof: Since α^u and α^x are at most 1, constraint (A2) will be satisfied whenever $\rho^u + \rho^x < q$, that is, whenever the total number of experienced workers is less than the number of jobs in the professional sector. We first restrict attention to this case and then argue that the optimal plan ensures $0 \leq \rho^u_t + \rho^x_t < q$ for all $t \geq 1$ starting at any initial condition (ρ^u_0, ρ^x_0).

The Bellman equation (A1) can be written as a functional equation

$$V = T(V),$$

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

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

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

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

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

Both of these are straightforward to verify. Define

$$\begin{aligned} \bar{T}(V; \alpha^u, \alpha^x, n) = {} & v^o + v_0^a(n) + \alpha^u \rho^u (v_1^a - v_0^a(n)) \\ & + \alpha^x \rho^x (v_1^p - v_0^a(n)) \\ & + \beta V((q - \rho^u \alpha^u - \rho^x \alpha^x)(1 - \phi(n)), \\ & \quad (q - \rho^u \alpha^u - \rho^x \alpha^x) \pi \phi(n)), \end{aligned}$$

that is, $\bar{T}(V)$ evaluates V for a fixed vector (α^u, α^x, n) rather than for the vector that maximizes the right-hand side of equation (A1) subject to constraints. Let $(\hat{\alpha}^u, \hat{\alpha}^x, \hat{n})$ be the values that solve the right-hand side of equation (A1) when $V = V_1$, which are functions of (ρ^u, ρ^x). If $V_2 \geq V_1$, then we then have

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

This establishes monotonicity. For discounting, observe that replacing V with $V + a$ will leave the arg max on the right-hand side of equation (A1) unchanged. Hence,

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

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

Next, we argue that V is linear in ρ^u and ρ^x over the set of all (ρ^u, ρ^x) for which $0 \leq \rho^u + \rho^x < q$. To prove this, it will be enough to show that if V is linear over this region, then $T(V)$ must be linear over this region as well.

In that case, the contraction mapping theorem tells us there exists a fixed point within the set of functions V that are linear over this region, ensuring that V is linear in this region.

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

$$V(\rho^u, \rho^x) = K_1 + K_2\rho^u + K_3\rho^x \quad \text{(A3)}$$

when $0 \le \rho^u + \rho^x < q$. Since $v_0^a(n)$ and $\phi(n)$ are both concave in n, the right-hand side of equation (A1) is concave in n when $V(\rho^u, \rho^x)$ is linear. Hence, for any value of α^x and α^u, the necessary and sufficient condition for $\hat{n}$ to be optimal is

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

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

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

It follows that the optimal $\hat{n}$ that maximizes equation (A1) is independent of ρ^u and ρ^x whenever $0 \le \rho^u + \rho^x < q$, although it will depend on the coefficients K_2 and K_3. Next, since the objective function above is linear in α^u and α^x, we can deduce that the following scheme is optimal:

$$\hat{\alpha}^u = \begin{cases} 1 & \text{if } v_1^a - v_0^a - \beta[\pi\phi(n)K_3 + (1 - \phi(n))K_2] \ge 0, \\ 0 & \text{if } v_1^a - v_0^a - \beta[\pi\phi(n)K_3 + (1 - \phi(n))K_2] < 0, \end{cases} \quad \text{(A4)}$$

$$\hat{\alpha}^x = \begin{cases} 1 & \text{if } v_1^p - v_0^a - \beta[\pi\phi(n)K_3 + (1 - \phi(n))K_2] \ge 0, \\ 0 & \text{if } v_1^p - v_0^a - \beta[\pi\phi(n)K_3 + (1 - \phi(n))K_2] < 0, \end{cases} \quad \text{(A5)}$$

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

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

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

$$\bar{K}_1 = v^o + v_0^a + \beta q[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2] + \beta K_1,$$
$$\bar{K}_2 = \hat{\alpha}^u(v_1^a - v_0^a) - \beta\hat{\alpha}^u[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2],$$
$$\bar{K}_3 = \hat{\alpha}^x(v_1^p - v_0^a) - \beta\hat{\alpha}^x[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2].$$

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

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

$$\rho_{t+1}^u + \rho_{t+1}^x = (q - \rho_t^u\hat{\alpha}^u - \rho_t^x\hat{\alpha}^x)(1-(1-\pi)\phi(\hat{n})). \quad \text{(A6)}$$

Note that both expressions on the right-hand side of equation (A6) are nonnegative. The first expression is nonnegative from constraint (A2), while the second is positive given our assumptions that the maximal effort level $\bar{n}$ is such that $\phi(\bar{n}) < 1$. It follows that $\rho_{t+1}^u + \rho_{t+1}^x \geq 0$. For a given (ρ_t^u, ρ_t^x), we consider two possible cases. First, suppose $\rho_t^u\hat{\alpha}^u + \rho_t^x\hat{\alpha}^x > 0$. In this case, the first term on the right-hand side of equation (A6) will be less than q, while the second term on the right-hand side of equation (A6) will be at most 1. Hence, $\rho_{t+1}^u + \rho_{t+1}^x < q$. Next, suppose $\rho_t^u\hat{\alpha}^u + \rho_t^x\hat{\alpha}^x = 0$. In this case, young associates will be employed at date t and asked to put in a positive amount of effort, in which case the optimal effort of young workers will satisfy $\hat{n}_t > 0$. In this case, the first term on the right-hand side of equation (A6) will be at most q, while the second term in equation (A6) will be strictly less than 1. In that case, we would once again have $\rho_{t+1}^u + \rho_{t+1}^x < q$. By induction, we can conclude that regardless of the initial condition (ρ_0^u, ρ_0^x), the optimal path will imply $\rho_t^u + \rho_t^x < q$ for all $t \geq 1$. QED

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

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

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

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

$$K_1 = (2-q)v^o + qv_0^a + \beta q[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2] + \beta K_1, \quad \text{(A7)}$$

$$K_2 = \hat{\alpha}^u(v_1^a - v_0^a) - \beta\hat{\alpha}^u(\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2), \quad \text{(A8)}$$

$$K_3 = \hat{\alpha}^x(v_1^p - v_0^a) - \beta\hat{\alpha}^x(\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2). \quad \text{(A9)}$$

Using equations (A8) and (A9) to solve for K_2 and K_3 yields

$$K_2 = \frac{\hat{\alpha}^u(v_1^a - v_0^a) - \beta\hat{\alpha}^u\hat{\alpha}^x\pi\phi(\hat{n})(v_1^p - v_1^a)}{1 + \beta\hat{\alpha}^u(1 - \phi(\hat{n})) + \beta\hat{\alpha}^x\pi\phi(\hat{n})}, \tag{A10}$$

$$K_3 = \frac{\hat{\alpha}^x(v_1^p - v_0^a) + \beta\hat{\alpha}^u\hat{\alpha}^x(1 - \phi(\hat{n}))(v_1^p - v_1^a)}{1 + \beta\hat{\alpha}^u(1 - \phi(\hat{n})) + \beta\hat{\alpha}^x\pi\phi(\hat{n})}. \tag{A11}$$

Note that both K_2 and K_3 are nonnegative. The first-order condition for $\hat{\alpha}^x$, in line with equation (A5), implies $\hat{\alpha}^x = 1$ whenever

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

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

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

Begin with K_2. Observe that

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

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

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

Next, consider K_3. Observe that

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

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

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

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

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

Proof: Consider the expression for K_1 as defined in equation (A7). This corresponds to the present discounted surplus the planner can expect to generate starting with no experienced workers, that is, $\rho^u = \rho^x = 0$. Since the planner always assigns $2 - q$ workers to the outside sector and they each produce a surplus of v^o, the present discounted value of surplus produced in the outside sector is just $(1 - \beta)^{-1}(2 - q)v^o$. Hence, the difference $K_1 - (1 - \beta)^{-1}(2 - q)v^o$ is equal to the net surplus the planner expects to generate in the professional sector when $\rho^u = \rho^x = 0$, that is, it is equal to the value of staffing all q positions in the professional sector with young workers, observing their output, and then staffing these positions optimally thereafter. Since information on worker types can be used to set their hours optimally, it follows that this value must be strictly larger than the surplus generated in the professional sector from staffing all q positions in the professional sector with young workers, ignoring any information that may be revealed about their quality and treating all such workers as if their ability was uncertain but acting optimally in using information thereafter. The latter yields a value of

$$qv_0^a + \beta\big(K_1 - (1 - \beta)^{-1}(2 - q)v^o + qK_2\big).$$

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

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

which after rearranging yields

$$(1-\beta)K_1 > (2-q)v^o + qv_0^a + \beta q K_2. \tag{A12}$$

From equation (A7) we know that K_1 satisfies

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

Rearranging this equation implies

$$\beta\phi(\hat{n})(\pi K_3 - K_2) = (1-\beta)K_1 - (2-q)v^o - qv_0^a - \beta q K_2. \tag{A13}$$

The right-hand side of equation (A13) is strictly positive given inequality (A12). Since it will always be optimal to have inexperienced workers put in some effort, $\phi(\hat{n}) > 0$. Hence, $\beta\phi(\hat{n})(\pi K_3 - K_2) > 0$. It follows that $\pi K_3 - K_2 > 0$. QED

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

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

$$c'(\hat{n}) = (1+\pi x) + \beta\phi'(\hat{n})(\pi K_3 - K_2). \tag{A14}$$

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

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

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

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

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

$$\hat{\alpha}^u = \begin{cases} 1 & \text{if } v_1^a - v_0^a - \beta[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2] > 0, \\ [0,1] & \text{if } v_1^a - v_0^a - \beta[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2] = 0, \\ 0 & \text{if } v_1^a - v_0^a - \beta[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2] < 0. \end{cases} \tag{A15}$$

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

$$\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2 = \frac{\pi\phi(\hat{n})(v_1^p - v_0^a) + \hat{\alpha}^u(1-\phi(\hat{n}))(v_1^a - v_0^a)}{1 + \beta\hat{\alpha}^u(1-\phi(\hat{n})) + \beta\pi\phi(\hat{n})}, \tag{A16}$$

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

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

which implies

$$\begin{aligned}(1 + \beta\pi\phi(\hat{n}))(v_1^a - v_0^a) &\geq \beta\pi\phi(\hat{n})(v_1^p - v_0^a),\\ (v_1^a - v_0^a) &\geq \beta\pi\phi(\hat{n})(v_1^p - v_1^a).\end{aligned} \tag{A17}$$

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

$$\hat{\alpha}^u = \begin{cases} 1 & \text{if } v_0^a(n) < v_1^a - \beta\pi\phi(\hat{n})(v_1^p - v_1^a),\\ [0,1] & \text{if } v_0^a(n) = v_1^a - \beta\pi\phi(\hat{n})(v_1^p - v_1^a),\\ 0 & \text{if } v_0^a(n) > v_1^a - \beta\pi\phi(\hat{n})(v_1^p - v_1^a),\end{cases} \tag{A18}$$

which proves the result. QED

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

Proof: Consider a constrained planner's problem in which α^u is given and the planner can only choose effort n, that is,

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

where

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

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

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

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

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

$$\frac{\partial K_1^*}{\partial n} = 0, \tag{A19}$$

$$\frac{\partial^2 K_1^*}{\partial n^2} < 0. \tag{A20}$$

To see how n^* varies with z^p, we can look at how it varies with v_1^p, given that the latter is monotonically increasing in z^p. Totally differentiate equation (A19) to obtain

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

Using the expressions for K_2^* and K_3^* from $V(\rho^u, \rho^x; \alpha^u)$, one can show[26] that

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

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

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

[26] We verify this using Mathematica. Code is available on request.

where

$$\zeta(n^*) = -c''(n^*)(1 + \beta\alpha^u(1 - \phi(n^*)) + \beta\pi\phi(n^*)) \\ + \beta\phi''(n^*)[\pi(v_1^p - v_0^a) - \alpha^u(v_1^a - v_0^a) + \pi\beta\alpha^u(v_1^p - v_1^a)]. \tag{A21}$$

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

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

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

By an analogous argument,

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

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

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

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

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

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

Finally, we know from (A18) that for any value of v_1^p, the optimal $\hat{\alpha}^u$ is either uniquely equal to 0, uniquely equal to 1, or else any value between 0 and 1 is optimal. When any value of α^u between 0 and 1 is optimal, the objective function in the unconstrained planner's problem is independent of α^u. Hence, the value of n we chose to maximize K_1 will be the same whether we set $\hat{\alpha}^u = 0$ or $\hat{\alpha}^u = 1$. This implies that the value of n^* that maximizes K_1^* when $\hat{\alpha}^u = 0$ is the same that maximizes K_1^* when $\hat{\alpha}^u = 1$, that is, $n^*(0) = n^*(1)$ whenever all values in [0, 1] are optimal for α^u.

Although the optimal $\hat{\alpha}^u$ in equation (A18) is a correspondence if v_1^p that can take on multiple values for some realizations of v_1^p, let us define a function $\tilde{\alpha}^u(v_1^p)$ that is equal to the optimal $\hat{\alpha}^u$ for any value of v_1^p where the $\hat{\alpha}^u(v_1^p)$ is unique. For any v_1^p such that $\hat{\alpha}^u$ contains the entire interval

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

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

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

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

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

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

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

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

Substituting this into $\zeta(n)$ in equation (A21) implies that

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

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

$$\frac{dn^*}{dv_1^p} = \frac{\beta\pi(1 + \beta\alpha^u)\phi'(n) > 0}{(1 + \beta\pi\phi(n) + \beta\alpha^u(1 - \phi(n)))[c''(n) - \beta\pi\phi''(n)(v_1^p - v_1^a)]}.$$

Differentiating this with respect to α^u yields

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

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

Since we know that $\hat{\alpha}^u$ switches at most once, we need to determine whether there is a switch as we increase v_1^a and v_1^p and 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 not 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 regard to v_1^a, although we impose a restriction that $z^a < w_0 <$

$z^a(1 + \pi x)$, the second inequality was imposed only 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 $z^a \rightarrow 1$ to obtain a boundary condition for $\hat{\alpha}^u$. In the limit as $z^a \rightarrow 1$, it is optimal to set $\alpha^u = 0$ and employ a young worker who has some option value than an experienced worker who does not. This is true even if $1 + \pi x < w_0$, as long as the production parameters are such that the planner always wants to employ some new workers in the professional sector, with the possible exception of the initial period. It follows that as we increase α^u, if there is a transition, it must be from $\alpha^u = 0$ to $\alpha^u = 1$. QED

Appendix B

Learning about Learning

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

B1. Environment

We structure our model to replicate our screening model described above as much as possible. Once again, time is infinite, and workers live for 2 periods. There are also two sectors: professional and outside. The marginal product of tasks performed in the outside sector is a constant, $w^o > 1$, that does not vary with worker skill or experience.

In the professional sector, total output is determined by the number of tasks performed and by how well workers of different skill levels sort to different tasks. We assume there are three types of tasks that can be performed in the professional sector: tasks 1, 2, and 3. Furthermore, we assume that workers may possess one of three skill levels: high, medium, or low.

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

Workers of all skill levels are qualified to perform task 1, and this task yields a marginal product of 1 regardless of the skill level of the worker who performs it. Task 2 yields a marginal product $z^a > w^o > 1$ if the worker has a medium or high skill level, but low-skilled workers are unable to perform this task. Task 3 yields a marginal product of $z^p > z^a$ if the worker has a high skill level, but a worker at the medium or low skill level is not qualified to perform this task.

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

If a new associate with learning efficiency λ performs n tasks, she acquires s skills, where $s = \lambda n$. If s exceeds a cutoff level $\bar{s}$, the worker has reached the high skill level and is able to perform any of the three tasks in the professional sector. We refer to such a worker as a high type, or type h. If s is less than or equal to $\bar{s}$ but still greater than some lower cutoff $\underline{s} < \bar{s}$, the worker has achieved the medium skill level and is able to perform tasks 1 and 2 but not task 3. This worker is a medium type, or type m. Finally, if s is less than or equal to $\underline{s}$, the worker remains at the low skill level—that is, type l—and given our assumption $w^o > 1$, her second-period productivity is greatest in the outside sector.

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

$$\phi_l(n) \equiv \Pr(\lambda n \leq \underline{s}) = F\left(\frac{\underline{s}}{n}\right),$$

$$\phi_m(n) \equiv \Pr(\lambda n \in (\underline{s}, \bar{s}]) = F\left(\frac{\bar{s}}{n}\right) - F\left(\frac{\underline{s}}{n}\right),$$

$$\phi_h(n) \equiv \Pr(\lambda n > \bar{s}) = 1 - F\left(\frac{\bar{s}}{n}\right).$$

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

To create a structure that parallels our screening model, we refer to task 3 as the partner task, task 2 as the senior associate task, and task 1 as the new associate task. We define the surplus associated with these positions, v_1^p, v_1^a, and $v_0^a(n)$, as before. That is,

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

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

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

$$\begin{aligned} V(\rho^m, \rho^h) = \max_{\alpha^h, \alpha^m, n} & (2 - q)v^o + qv_0^a(n) + \alpha^h \rho^h (v_1^p - v_0^a(n)) \\ & + \alpha^m \rho^m (v_1^a - v_0^a(n)) + \beta V(\rho_{+1}^m, \rho_{+1}^h), \end{aligned} \tag{B1}$$

where

$$\rho_{+1}^m = (q - \rho^m \alpha^m - \rho^h \alpha^h)\phi_m(n), \tag{B2}$$

$$\rho_{+1}^h = (q - \rho^m \alpha^m - \rho^h \alpha^h)\phi_h(n), \tag{B3}$$

and the constraint

$$q - \alpha^m \rho^m - \alpha^h \rho^h \geq 0. \tag{B4}$$

The state variables ρ^m and ρ^h are analogous to ρ^u and ρ^x in the original screening model. The only features that do not have exact parallels in our screening model are in the laws of motion (B2) and (B3). In particular, in the screening model the analogs to $\phi_h(n)$ and $\phi_m(n)$ were required to satisfy $\phi_h(n) = \pi(1 - \phi_m(n))$, whereas now there is no analogous restriction. The counterpart to the restriction that $\phi'(n) > 0$ is that now $\phi_h'(n) > 0$. In the screening model, this would have implied $\phi_m'(n) < 0$, but this need not be the case in the present model, where the sign of $\phi_m'(n)$ is generally ambiguous. In what follows, we show that if $\phi_m'(n) \leq 0$, our learning-by-doing model yields results that are parallel to all of the results from our screening model. We also analyze the case where $\phi_m'(n) > 0$. In this case, we can quickly establish some results that parallel those from our screening model, but we need to impose an additional assumption to establish results that parallel all of the findings in Sections IV and V.

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

B2. Results That Are Independent of the Sign of ϕ_m'

Applying the same arguments as in the proof of claim 1, we can confirm that the Bellman equation for the planner's problem is once again linear in ρ^m and ρ^b whenever $0 \le \rho^m + \rho^b < q$, that is,

$$V(\rho^m, \rho^b) = K_1 + K_m\rho^m + K_b\rho^b.$$

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

$$\begin{aligned} K_1 &= (2-q)v^o + v_0^a(\hat{n}) + q\beta[\phi_m(\hat{n})K_m + \phi_b(\hat{n})K_b] + \beta K_1, \\ K_m &= \hat{\alpha}^m(v_1^a - v_0^a) - \beta\hat{\alpha}^m[\phi_m(\hat{n})K_m + \phi_b(\hat{n})K_b], \\ K_b &= \hat{\alpha}^b(v_1^p - v_0^a) - \beta\hat{\alpha}^b[\phi_m(\hat{n})K_m + \phi_b(\hat{n})K_b]. \end{aligned}$$

Solving this system yields

$$K_m = \frac{\hat{\alpha}^m(v_1^a - v_0^a) - \beta\hat{\alpha}^m\hat{\alpha}^b\phi_b(\hat{n})(v_1^p - v_1^a)}{1 + \beta\left[\hat{\alpha}^m\phi_m(\hat{n}) + \hat{\alpha}^b\phi_b(\hat{n})\right]}, \tag{B5}$$

$$K_b = \frac{\hat{\alpha}^b(v_1^p - v_0^a) + \beta\hat{\alpha}^b\hat{\alpha}^m\phi_m(\hat{n})(v_1^p - v_1^a)}{1 + \beta\left[\hat{\alpha}^m\phi_m(\hat{n}) + \hat{\alpha}^b\phi_b(\hat{n})\right]}. \tag{B6}$$

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

$$\phi_b''(n)K_b + \phi_m''(n)K_m - c''(n) < 0.$$

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

$$c'(\hat{n}) = 1 + \beta(\phi_b'(\hat{n})K_b + \phi_m'(\hat{n})K_m).$$

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

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

$$\hat{\alpha}^b = \begin{cases} 1 & \text{if } v_1^p - v_0^a - \beta[\phi_m(\hat{n})K_m + \phi_b(\hat{n})K_b] \ge 0, \\ 0 & \text{if } v_1^p - v_0^a - \beta[\phi_m(\hat{n})K_m + \phi_b(\hat{n})K_b] < 0. \end{cases}$$

As in the previous case, the fact that employment in the professional sector is equal to q means that there will be at most q inexperienced associates at any period $t \ge 1$. Since some of them will fail to master enough skills to perform new jobs, it follows that $\rho_{t+1}^m + \rho_{t+1}^h < q$. Once again, then, constraint (B4) will not bind beyond the initial period.

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

$$v_1^p - v_0^a \geq \beta[\phi_m(\hat{n})K_m + \phi_h(\hat{n})K_h].$$

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

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

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

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

Next, consider K_h. Observe that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

which implies

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

as desired.

B3. Results for the Case Where $\phi'_m \le 0$

We begin with the analog to proposition 3. As we did in the screening model, let $n^*(\alpha^m)$ denote the level of effort that solves the planner's problem for a given value of α^m. It must satisfy the first-order necessary condition

$$\frac{\partial K_1^*}{\partial n} = 0 \tag{B7}$$

as well as the second-order necessary condition

$$\frac{\partial^2 K_1^*}{\partial n^2} < 0. \tag{B8}$$

Totally differentiating the first-order condition yields

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

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

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

When $\phi'_m \le 0$, this expression is positive. Next, we have

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

where

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

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

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

Since $\zeta < 0$ at the optimum, it follows from the second-order condition that $(dn^*/dv_1^p) > 0$. Analogously, we have

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

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

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

$$= -\beta\alpha^m \frac{\phi'_m + \beta[\phi'_m\phi_h - \phi'_h\phi_m]}{\zeta} \le 0.$$

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

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

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

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

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

and so

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

Differentiating with respect to α^m yields

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

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

B4. Results for the Case Where $\phi'_m > 0$

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

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

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

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

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

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

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

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

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

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

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

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

$$\frac{dn^*}{dv_1^a} = -\beta\alpha^m\frac{\phi'_m - \beta[\phi'_h\phi_m - \phi'_m\phi_h]}{\zeta},$$

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

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

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

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

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

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

which can be either positive or negative, while

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

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

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

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

Appendix C

Endogenous Sector Size and Decentralization

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

C1. Planner's Problem with Endogenous Sector Size

When the number of jobs q is endogenous, the planner's problem is given by

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

subject to

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

as well as the constraint

$$q - \rho_t^u\alpha^u - \rho_t^x\alpha^x \geq 0. \tag{C1}$$

The problem described above is similar to our original planning problem except that rather than treating q as given, the planner now chooses q in addition to α^x, α^u, and n_0^u and incurs a cost $\int_0^q \kappa(x)\,dx$ that depends on the number of jobs created in the professional sector.

We begin by ignoring constraint (C1) and analyzing an unconstrained planning problem that does not involve this constraint. We can then verify whether the solution to the unconstrained problem satisfies this constraint. Let ξ_0^x denote the multiplier on the constraint that $\alpha^x \geq 0$ and ξ_1^x denote the multiplier on the constraint that $\alpha^x \leq 1$. Likewise, let ξ_0^u denote the multiplier on the constraint that $\alpha^u \geq 0$ and ξ_1^u denote the multiplier on the constraint that $\alpha^u \leq 1$. The first-order conditions of the maximization problem above with respect to n, α^x, α^u, and q are given by

$$n:(1+\pi x) - c'(n) - \beta\phi'(n)\left[\pi\frac{\partial V}{\partial \rho^x} - \frac{\partial V}{\partial \rho^u}\right] = 0, \tag{C2}$$

$$\alpha^x:\rho^x\left[(v_1^p - v_0^a) - \beta\left(\pi\phi(n)\frac{\partial V}{\partial \rho^x} - (1-\phi(n))\frac{\partial V}{\partial \rho^u}\right)\right] = \xi_1^x - \xi_0^x, \tag{C3}$$

$$\alpha^u:\rho^u\left[(v_1^a - v_0^a) - \beta\left(\pi\phi(n)\frac{\partial V}{\partial \rho^x} - (1-\phi(n))\frac{\partial V}{\partial \rho^u}\right)\right] = \xi_1^u - \xi_0^u, \tag{C4}$$

$$q: -v^o + v_0^a - \kappa(q) + \beta\left[\pi\phi(n)\frac{\partial V}{\partial \rho^x} + (1-\phi(n))\frac{\partial V}{\partial \rho^u}\right] = 0. \tag{C5}$$

Define $\mathbb{U}$ as the set of values (ρ^u, ρ^x) for which the optimal solution to the unconstrained problem implies $\hat{q} - \rho_t^u\hat{\alpha}^u - \rho_t^x\hat{\alpha}^x > 0$. At such values, constraint (C1) will not be binding, so within this set the optimal plan to the constrained planner's problem coincides with the unconstrained problem.

As in the case where q was fixed, we first argue that the function V is linear in (ρ^u, ρ^x) over the set $\mathbb{U}$. This is because the operator T implied by the Bellman equation described above maps functions V that are linear over the set $\mathbb{U}$ into functions $T(V)$ that are linear over the set $\mathbb{U}$. To see this, let us once again denote the optimal values with a hat. Equation (C2) implies that the optimal $\hat{n}$ will not vary with (ρ^u, ρ^x) in the set $\mathbb{U}$ given that V is linear over this set, meaning that $\partial V/\partial \rho_t^x$ and $\partial V/\partial \rho_t^u$ are constant. Note here that the argument relies on the optimal solution satisfying the condition that $\hat{q} - \rho_t^u\hat{\alpha}^u - \rho_t^x\hat{\alpha}^x > 0$. From equations (C3) and (C4), we can then conclude that the optimal $\hat{\alpha}^u$ and $\hat{\alpha}^x$ do not vary with (ρ^u, ρ^x) in the set $\mathbb{U}$. In particular, when $V(\rho^u, \rho^x)$ is linear and the derivatives in equation (C3) and (C4) are independent of (ρ^u, ρ^x), the derivative with respect to α^x and α^u will either be positive, zero, or negative for all $(\rho^u, \rho^x) \in \mathbb{U}$, in which case the optimal $\hat{\alpha}^u$ and $\hat{\alpha}^x$ will be the same regardless of the value of the multipliers ξ_1^x, ξ_1^u, ξ_0^x, and ξ_0^u. Finally, from equation (C5) we can say that the optimal $\hat{q}$ does not vary with (ρ^u, ρ^x) in the set $\mathbb{U}$. Given that the optimal vector $(\hat{\alpha}^x, \hat{\alpha}^u, \hat{n}, \hat{q})$ is constant over $\mathbb{U}$, the function $T(V)$ must be linear in (ρ^u, ρ^x) over the set $\mathbb{U}$ as well. Since the Bellman equation V is the unique fixed point that solves $V = T(V)$, it follows that the Bellman equation is linear in (ρ^u, ρ^x) over the set $\mathbb{U}$, that is,

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

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

Since the optimal $\hat{q}$ does not vary with (ρ^u, ρ^x) in the set $\mathbb{U}$, we can denote this common value by q^*. From equation (C5), we know that q^* solves

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

Recall that K_2 and K_3 are defined independently of q, so the above yields a closed-form expression for q^*. We can now argue that the set $\mathbb{U}$ is nonempty. Consider the set $\Omega \equiv \{(\rho^u, \rho^x) \in \mathbb{R}_+^2 : \rho^u + \rho^x < q^*\}$. Since $q^* > 0$, the set Ω is nonempty. Moreover, since α^x and α^u are at most 1, it follows that $\Omega \subset \mathbb{U}$, ensuring that the latter is nonempty. That is, ignoring constraint (C1) yields a unique optimal sector size q^* the planner would prefer, and as long as the total number of experienced workers is below q^*, constraint (C1) will not bind.

Finally, we argue that under the optimal plan, $(\rho_t^u, \rho_t^x) \in \mathbb{U}$ for all $t \geq 1$, that is, constraint (C1) will cease to bind after at most 1 period. At date 0, we can have $(\rho_0^u, \rho_0^x) \in \mathbb{U}$ or $(\rho_0^u, \rho_0^x) \notin \mathbb{U}$. If $(\rho_0^u, \rho_0^x) \in \mathbb{U}$, then the planner will choose $q = q^*$. In this case, we know from the law of motion that

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

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

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

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

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

That is, if $(\rho_t^u, \rho_t^x) \in \mathbb{U}$, then $(\rho_{t+1}^u, \rho_{t+1}^x) \in \mathbb{U}$. This leaves us with the case $(\rho_0^u, \rho_0^x) \notin \mathbb{U}$. In this case, constraint (C1) is strictly binding, meaning

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

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

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

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

Finally, we establish the following comparative static result:

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

Proof: From the first-order condition for the planner's problem, we have

$$\begin{aligned}\kappa(\hat{q}) &= v_0^a - v^o + \beta\left[\pi\phi(\hat{n})\frac{\partial V}{\partial \rho^x} + (1-\phi(\hat{n}))\frac{\partial V}{\partial \rho^u}\right] \\ &= v_0^a - v^o + \beta[\pi\phi(\hat{n})K_3 + (1-\phi(\hat{n}))K_2].\end{aligned} \tag{C6}$$

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

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

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

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

where we have replaced K_2^* and K_3^* with K_2 and K_3, the value for K_2^* and K_3^* when we set α^u to the optimal value, $\hat{\alpha}^u$. Hence, at the optimum, changing $\hat{n}$ will have no effect on the right-hand side of equation (C6). Thus, at the optimal allocation, the right-hand side of equation (C6) must be increasing in v_1^p and weakly increasing in v_1^a. Since $\kappa(q)$ is assumed to be increasing in q, the optimal $\hat{q}$ will be higher. QED

C2. Decentralization

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

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

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

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

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

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

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

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

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

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

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

But from the first-order condition (C5), we have

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

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

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

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

Next, we verify that experienced workers whose type is uncertain work in the professional sector if and only if the planner would assign such workers to the professional sector, that is, if and only if $\hat{\alpha}^u = 1$. Again, using first-order condition (C5), we know that

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

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

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

Moreover, we know from proposition 2 that $\hat{\alpha}^u = 0$ if and only if

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

Rearranging this inequality implies

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

and so

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

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

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

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

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

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

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

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

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

$$c'(\tilde{n}) = 1 + \pi x + \beta\phi'(\tilde{n})[\pi v_1^p - \tilde{\alpha}^u v_1^a + (\tilde{\alpha}^u - \pi)(v^o + \kappa(\hat{q}))]. \quad \text{(C7)}$$

From the planner's first-order condition (C5), substituting in for K_2 and K_3 and the fact that $\tilde{\alpha}^u = \hat{\alpha}^u$, we can rewrite the first-order condition (C7) as

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

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

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

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

which on rearranging implies

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

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

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

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

Since $\hat{n}$ is optimal, we know that the right-hand side above is higher at $\hat{n}$ than at the static optimum and the value at the static optimum, which in turn exceeds the value of working in the professional sector at the static optimal level for only 1 period. This further implies that workers who start in the professional sector are strictly worse off in their first period than workers who start in the outside sector, a result we refer to in the text.

Appendix D

Data

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

D1. Survey of Law Firm Economics (SLFE)

Table 1 uses data from the SLFE. This survey is conducted annually by ALM Legal Intelligence. We obtained electronic versions of the data for the eight surveys conducted between 2007 and 2014.

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

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

Table 1 contains data on three lawyer-level variables: hours billed, hourly billing rate, and total compensation. We define total compensation as the sum of three measures collected by ALM: salary, bonus, and benefits. We express all monetary variables in 2011 dollars using the consumer price index for all urban consumers (CPI-U) as our inflation measure. This facilitates comparisons with the After the JD data in tables 2 and 3.

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

The firm-level data report totals for each type of lawyer employed in each firm. We sum these totals to create our firm size variable. Our measure of firm size is the sum of full-time equivalent lawyers in each firm.

We adopted several rules for cleaning the data. We corrected several obvious coding errors in the year-barred variable.[27] When calculating compensation variables, we treat total compensation as missing if a lawyer (*a*) reports a salary below $10,000 or above $5,000,000, (*b*) reports a bonus greater than $5,000,000, (*c*) reports a benefit amount less than $0 or greater than $100,000, or (*d*) reports less than 0 or more than 3,000 billable hours. We also code billing rates greater than $1,200 per hour as missing.

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

D2. After the JD (AJD) Survey

Tables 2 and 3 use data from the AJD survey, conducted by the American Bar Association (ABA) in three waves from 2002 to 2012. This longitudinal survey followed a stratified random sample of lawyers who were first admitted to the bar around 2000. The first stage of the two-stage sampling pro-

[27] The full list of changes is as follows: 205 became 2005, 208 became 2008, 1190 became 1990, and 1194 and 1794 became 1994. Additionally, any year less than 100 became that year plus 1900. There were no entries between 0 and 15, so it was appropriate to add 1900 in all cases rather than 2000.

cess divided the country into 18 strata based on the number of new lawyers in each area. In the second stage, researchers chose one primary sampling unit from each strata. These sampling units are local markets for legal services. The largest are the four "major" legal markets: Chicago, New York, Los Angeles, and Washington, DC. These markets contain more than 2,000 new lawyers. Small states make up some of the smaller sampling units.

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

We restrict the samples used in tables 2 and 3 to respondents who (1) responded to both the wave 1 survey and the wave 3 survey and (2) passed the bar in or after 1998. We divide our sample of lawyers into five categories based on the position that they reported in wave 3. Our categories include four positions in private law firms that are not run by solo practitioners. These positions are partner, nonequity partner, associate, and of counsel (counsel). We group all other lawyers in an "other" category. This category contains persons who no longer work in a private law firm, solo practitioners, and a small number of contract or staff attorneys.

The ALM survey gave descriptions of different law firm positions in the survey instrument. The AJD does not provide definitions of the four positions we highlight. We argue in Section VI that most nonequity partners, especially those who have roughly 10 years of experience, are still trying to earn promotion to full equity partner and in some cases already function as partners in their interactions with clients. The AJD data provide support for this claim. All respondents to wave 3, not just the ones we select for our samples in tables 2 and 3, provide a retrospective employment history. In wave 3, 1,472 lawyers report that they began their careers as associates in private law firms.[28] Only 92 of these lawyers report making partner in their original firm, and 18 of these lawyers report that they made partner after working as a nonequity partner in their original firm. Another 77 were pro-

[28] The data for tables 2 and 3 come from lawyers who responded to both the wave 1 survey and the wave 3 survey. This accounts for the significantly smaller sample sizes.

moted to nonequity partner in their original firm but had not made partner in the firm by wave 3. Of these, 22 left the firm before wave 3, and 55 remained.

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

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

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

The AJD does not provide weights that adjust for differential attrition between waves 1 and 3. We did create versions of tables 2 and 3 using the wave 1 sampling weights and found patterns that are quite similar to those in our unweighted analyses.

References

Akerlof, George. 1976. The economics of caste and of the rat race and other woeful tales. *Quarterly Journal of Economics* 90, no. 4:599–617.

Almer, Elizabeth Dreike. 2004. AICPA work/life and women's initiatives 2004 research: Executive summary.

Barlevy, Gadi, and Derek Neal. 2016. Allocating effort and talent in professional labor markets. Federal Reserve Bank of Chicago Working Paper no. 2016-03.

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

Bertrand, Marianne, Claudia Goldin, and Lawrence Katz. 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economics Journal: Applied Economics* 2:228–55.

Bonatti, Alessandro, and Johaness Horner. 2017. Career concerns with exponential learning. *Theoretical Economics* 12, no. 1:425–75.

Bose, Gautam, and Kevin Lang. 2017. Monitoring for worker quality. *Journal of Labor Economics* 35, no. 3:755–85.

Demougin, Dominique, and Aloysius Siow. 1994. Careers in ongoing hierarchies. *American Economic Review* 84, no. 5:1261–77.

Flom, Barbara, and Stephanie A. Scharf. 2011. Report of the Sixth Annual National Survey on Retention and Promotion of Women in Law Firms. Chicago: National Association of Women Lawyers and the NAWL Foundation.

Galanter, Marc, and William D. Henderson. 2008. The elastic tournament: A second transformation of the big law firm. *Stanford Law Review* 60, no. 6:1867–930.

Galanter, Marc, and Thomas Palay. 1991. *Tournament of lawyers: The transformation of the big law firm*. Chicago: University of Chicago Press.

Ghosh, Suman, and Michael Waldman. 2010. Standard promotion practices versus up-or-out contracts. *RAND Journal of Economics* 41, no. 2:301–25.

Gicheva, Dora. 2013. Working long hours and early career outcomes in the high-end labor market. *Journal of Labor Economics* 31, no. 4:785–824.

Gilson, Ronald J., and Robert H. Mnookin. 1985. Sharing among the human capitalists: An economic inquiry into the corporate law firm and how partners split profits. *Stanford Law Review* 37, no. 2:313–92.

———. 1989. Coming of age in a corporate law firm: The economics of associate career patterns. *Stanford Law Review* 41, no. 3:567–95.

Gorman, Elizabeth H. 1999. Moving away from up-or-out: Determinants of permanent employment in law firms. *Law and Society Review* 33, no. 3:637–66.

Greenhouse, Steven. 2011. Flex time flourishes in accounting industry. *New York Times*, January 7.

Holmstrom, Bengt. 1999. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies* 66, no. 1:169–82.

Kahn, Charles, and Gur Huberman. 1988. Two-sided uncertainty and up-or-out contracts. *Journal of Labor Economics* 6, no. 4:423–44.

Landers, Renee M., James B. Rebitzer, and Lowell J. Taylor. 1996. Rat race redux: Adverse selection in the determination of work hours in law firms. *American Economic Review* 86, no. 3:329–48.

Lewison, John. 2006. The work/life balance sheet so far. *Journal of Accountancy* 202, no. 2:45–49.

O'Flaherty, Brendan, and Aloysius Siow. 1992. On the job screening, up or out rules, and firm growth. *Canadian Journal of Economics* 25, no. 2:346–68.

Prendergast, Canice. 1993. The role of promotion in inducing specific human capital acquisition. *Quarterly Journal of Economics* 108, no. 2:523–34.

Rebitzer, James, and Lowell Taylor. 2007. When knowledge is an asset: Explaining the organizational structure of large law firms. *Journal of Labor Economics* 25, no. 2:201–29.

Richmond, Douglas. 2010. The partnership paradigm and law firm non-equity partners. *Kansas Law Review* 58:507–51.

Waldman, Michael. 1990. Up-or-out contracts: A signaling perspective. *Journal of Labor Economics* 8, no. 2:230–50.

Wilkins, David B., and G. Mitu Gulati. 1998. Reconceiving the tournament of lawyers: Tracking, seeding, and information control in the internal labor markets of elite law firms. *Virginia Law Review* 84:1581–681.