Managing Careers in Organizations*

Rongzhu Ke Jin Li Michael Powell

December 15, 2014

Abstract

A firm's organizational structure imposes constraints on its ability to use promotion-based incentive systems. The main contribution of this paper is to develop a framework for identifying these constraints and exploring their consequences. We show that firms manage workers' careers by putting in place personnel policies that optimally resemble an internal labor market. Firms may adopt forced-turnover policies in order to keep lines of advancement open. In addition, they may alter their organizational structures to relax these constraints. This gives rise to a trade-off between incentive provision at the worker level and productive efficiency at the firm level. Our framework generates novel testable implications that connect firm-level characteristics with workers' careers in ways that are consistent with a rich set of empirical findings.

*Rongzhu Ke: Department of Economics, Chinese University of Hong Kong. E-mail: rzke@cuhk.edu.hk. Jin Li: Department of Management and Strategy, Kellogg School of Management, Northwestern University. E-mail: jin-li@kellogg.northwestern.edu. Michael Powell: Department of Management and Strategy, Kellogg School of Management, Northwestern University. E-mail: mike-powell@kellogg.northwestern.edu. We thank Chen Cheng and Can Urgun for excellent research assistance. We are grateful to Daron Acemoglu, Ricardo Alonso, Pierre Azoulay, Dan Barron, Tim Bond, Jimmy Chan, Yuk-fai Fong, Bob Gibbons, Bengt Holmstrom, Hideshi Itoh, Ben Jones, Kevin Lang, Danielle Li, Niko Matouschek, David Matsa, Giuseppe Moscarini, Arijit Mukherjee, Paul Oyer, Nicola Persico, Ivan Png, Jim Rebitzer, Armin Schmutzler, Xianwen Shi, Aloysius Siow, Michael Song, Wing Suen, Mike Waldman,Yanhui Wu, Jano Zabojnik, and seminar participants at CEPR, Clemson, CUHK, Fudan, Hitotsubashi, HKU, HKUST, Kyoto, MIT, Nagoya, NBER, NUS, Purdue, Queens, Regensburg, SHUFE, SOLE, UCLA, USC Marshall, and UToronto (CEPA Brown Bag) for helpful conversations and suggestions.

1 Introduction

Alfred P. Sloan's description of the firm as "a pyramid of opportunities" captures two fundamental elements of human resource management in firms. The first is that firms are able to use promotion opportunities to motivate and retain their workers. The second is that the promotion policies firms can put in place are constrained by the realities of their org chart. Common problems firms face arise from a mismatch between these considerations. If the firm insists on a particular org chart, then it cannot promote all the workers it may like to, and as Peter Cappelli (2008) observed, "Frustration with advancement opportunities is among the most important factors pushing individuals to leave for jobs elsewhere." But overemphasizing advancement opportunities while neglecting production results in top-heavy organizations that fail to keep up with their competition.

To date, the economics literature has mostly focused on the benefits of firms providing promotion opportunities. The constraints firms face in doing so have been beyond the scope of leading models of internal labor markets, which either treat workers' careers independently (Harris and Holmstrom, 1982; Waldman, 1984; MacLeod and Malcomson, 1988; Gibbons and Waldman, 1999a) or take promotion opportunities as given (Lazear and Rosen, 1981). As a result, the ways in which firms manage their workers' careers in order to cope with these constraints have received little attention.[1] In this paper, we develop a parsimonious framework in which promotion opportunities motivate workers, but firms face a "budget constraint" on promotion opportunities. Firms manage their workers' careers subject to this constraint, and they can alter their organizational structure in order to relax this constraint. Our framework generates novel testable implications that connect firm-level characteristics with workers' careers in ways that are consistent with many recent empirical findings.

Our starting point is a dynamic moral-hazard model in which promotions arise as an optimal way to motivate workers. We assume that contracting imperfections limit transfers and therefore require the firm to provide workers with rents to motivate them. Using familiar reasoning from dynamic moral-hazard models, we show that these rents should be backloaded in a worker's career. If production requires multiple activities, then activities requiring more rents are optimally performed by workers farther along in their careers: promotions arise as a way of optimally reusing rents.

Having established that promotions are the optimal way to motivate workers, we turn to the heart of the problem and highlight the tension that arises between productive efficiency and incentive provision. Under standard production-efficiency conditions, the firm would choose the number of workers performing each activity in order to equate the marginal revenue of their production to

[1] A notable exception is Waldman's (2003) model showing that such constraints give rise to insider bias in hiring.

the wages they are paid. But the resulting number of positions may provide inadequate promotion opportunities for the firm's workers. A tension therefore arises between using promotions to provide incentives for a given worker and using activity assignment for productive efficiency. In our model, firms optimally choose their personnel policies and the number of workers performing each activity in order to balance incentive provision and productive efficiency. As a consequence, firm-level characteristics, such as its technology, drive the firm's choices and therefore determine the career paths of workers.

Model. Our model builds upon Shapiro and Stiglitz's (1984) efficiency-wage model by allowing for multiple activities within a single firm. Homogeneous workers privately choose whether to work or shirk, and the firm can motivate workers by committing to a wage that is tied to the activity, coupled with the threat of firing workers who are caught shirking. Each firm has two types of activities that have to be performed, and each worker can perform a single type of activity in each period. The two activities differ in the level of rents required to provide motivation, because one activity (the high-rent activity) is either more onerous or more difficult to monitor than the other (the low-rent activity). The firm's output and therefore its revenues depend on how many workers perform each activity, and the firm maximizes its steady-state profits.

To do so, the firm has to choose the number of positions that will be available for workers performing each activity. In addition, the firm chooses a bundle of personnel policies. How many workers should the firm hire for each activity each period? Should the firm retain its incumbent workers? If so, what activity should they perform next period? What wage should be associated with each activity? The firm's personnel policies are limited by two key constraints. Workers have to be motivated to exert effort in each activity. That is, each worker's incentive-compatibility constraint must be satisfied. Additionally, for the firm to be in a steady state, a flow constraint must be satisfied: the number of incumbents and new hires who flow into each activity must equal the number of workers who flow out of that activity in each period.

Results. Optimal personnel policies reflect the tension between worker motivation and productive efficiency and therefore exhibit two sets of features, which are consistent with many stylized facts. The first set of features arises because rewards are optimally backloaded in a worker's career and would arise in a model with a single worker. This set of features resembles an internal labor market. The low-rent activity is performed in the bottom job, which serves as a port of entry (Doeringer and Piore, 1971). Workers remain in the bottom job until they are promoted to the top job, in which they perform the high-rent activity. Once in the top job, workers are never demoted (Baker, Gibbs, and Holmstrom 1994a). As a result, a well-defined career path emerges, and it plays the role of workers' "trust funds" (Akerlof and Katz, 1989): workers in the bottom job receive zero rents, effectively posting a bond by beginning employment in the bottom job. Their pay is backloaded through a high wage in the top job, which in turn is high enough to motivate effort in the high-rent activity. A worker's wages therefore increase upon promotion.

The second set of features of optimal personnel policies arise precisely because firms have multiple workers. When a worker departs from the firm, his position can be reallocated to another worker. Worker turnover therefore can expand promotion opportunities, providing a reason for why the firm might want to put in place forced-turnover policies such as mandatory-retirement programs. If the promotion prospects created solely from voluntary turnover at the top are insufficient for motivating workers at the bottom, the firm optimally forces a fraction of the workers at the top to leave the firm in every period. Viewed in isolation, adopting forced-turnover policies is a bad idea, since doing so reduces the expected rents of workers at the top, which would violate their incentive-compatibility constraint. However, forced-turnover policies are optimally complemented with more generous compensation for workers at the top as well as a more generous promotion policy for workers at the bottom.

In addition to choosing the personnel policies, the firm may expand promotion opportunities by altering the number of positions away from what would be productively efficient. Creating an additional top position expands the opportunities available for those at the bottom and therefore confers a benefit to the firm in addition to marginal revenue. In contrast, creating an additional bottom position reduces the promotion prospects of those at the bottom, and therefore the benefit of doing so is less than the marginal revenue the position creates. For both of these reasons, the firm's organizational span—the ratio of the number of positions at the bottom to the number of positions at the top—is optimally lower than would be productively efficient for the wages it pays.

External Factors that Affect Workers' Careers. Our model suggest two novel features of optimal personnel policies: workers' careers are optimally interlinked, and firm-level factors determine how the firm trades off productive efficiency with incentive provision. These features have fresh implications for how factors that are external to a given worker can affect a worker's career nonetheless.

Since workers' careers are optimally interlinked, a firm's demographics affects the careers of *all* its workers. As we mentioned above, firms may optimally create promotion opportunities for younger workers by putting in place forced-turnover policies that are targeted at older workers. At the economy-wide level, many countries have expanded the generosity of government retirement programs in order to encourage turnover of older workers and create opportunities for younger workers, but such policies have often had the opposite effects (see Gruber and Wise, 2010, for many studies documenting these results). Our model predicts that, for some parameter ranges, such government policies may indeed negatively affect younger workers' employment prospects, precisely because firms optimally alter their personnel policies and the number of workers they demand in response. We therefore provide an organizational explanation for the findings documented by the studies contained in Gruber and Wise (2010).

Since workers' careers are optimally managed at the firm level, they are determined in part by a firm's characteristics, such as its size. A large empirical literature has established that larger firms tend to pay higher wages on average (see Oi and Idson (1999) for a review of the firm size–wage relationship). Our model predicts that if larger firms have larger spans, which empirically they do, they will put in place different personnel policies. These different policies involve higher wages at all levels of the firm, so our model generates a firm size–wage relationship. Our model goes beyond providing yet another explanation for this relationship, though, and yields additional predictions about the relationship among firm size, personnel policies, and worker career dynamics. For example, larger firms will tend to have lower promotion rates, and they will have larger insider biases in hiring into higher-level positions. These policies are consistent with more recent empirical findings (DeVaro and Morita, 2013; Bond, 2014). More importantly, they underscore the general point that two identical workers may have different career experiences depending on the characteristics of their employers, precisely because different employers will optimally put in place different personnel policies.

To conclude, we note that a key result in our main model is that the firm has to offer workers rents in order to motivate them. These rents arise, because wages are tied to the activities workers perform, which prevents the firm from using contingent bonuses in order to provide motivation and low ex-ante wages in order to fully extract workers' expected bonuses. In order to optimally motivate workers while minimizing the rents the firm needs to ofer them, the firm manages its workers' careers by choosing its personnel policies and organizational structure. From an applied modeling perspective, the assumption that wages are tied to activities is stronger than we need in order to ensure that the firm has to provide its workers with rents.

In an extension, we consider a setting in which the firm is able to offer a general class of contracts involving history-dependent activity-assignment and bonus schemes, and workers are protected by a limited-liability constraint. Optimal contracting in this setting preserves most of the features of optimal personnel policies in the main model: internal labor markets arise, the activity requiring greater incentive rents becomes the top job, and the structure of optimal contracts depends critically on whether the promotion opportunities created by voluntary turnover among workers performing the high-rent activity are sufficient for motivating workers performing the low-rent activity.

In contrast to the main model, when such promotion opportunities are insufficient for motivating workers performing the low-rent activity, firms do not put in place forced-turnover policies. Rather, the firm rewards workers in the bottom job with performance bonuses. These differences suggest that forced-turnover policies are more appealing in environments in which the use of pay-for-performance contracts is limited. This difference accords with our broader findings that personnel policies are interconnected—the availability of pay-for-performance affects not only the firm's compensation policy, but also its choice of retention policies. Finally, a notable feature of the extension is that firms continue to use promotions to motivate workers, and they use bonuses for bottom workers only when promotions alone are not sufficient for providing motivation. Our model therefore sheds light on the Baker, Jensen, and Murphy (1988) puzzle that promotions are used as the predominant way of providing incentives, even in situations where they ought to be dominated by performance pay.

Literature Review This paper contributes to the literature on internal labor markets (see Gibbons (1997), Gibbons and Waldman (1999b), Lazear (1999), Lazear and Oyer (2013), and Waldman (2013) for reviews of the theory on and evidence for internal labor markets). In particular, our focus on promotions is related to the extensive literature on promotions as tournaments, starting with the seminal work of Lazear and Rosen (1981). Relative to these papers, our emphasis on the ongoing nature of careers in firms imposes restrictions on how prizes are optimally structured. In particular, firms have to ensure that workers have the incentives to exert effort at all times. This is why, in our model, promotions are accompanied with wage increases rather than one-time bonuses—such "efficiency wages" serve to motivate the promoted workers. Workers who are not promoted remain in the running for future promotions, and the firm has to make sure that such promotion opportunities arise frequently enough to keep these workers motivated. The way they do so is by choosing hiring and retention policies that keep the line of advancement open.

Our model focuses on how firms structure their internal labor markets in order to provide incentives to workers (Malcomson, 1984; MacLeod and Malcomson, 1988; Prendergast, 1993; Zabojnik and Bernhardt, 2001; Waldman, 2003; Camara and Bernhardt, 2009; Krakel and Schottner, 2012; Auriol, Friebel, and von Bieberstein, 2013). In contrast to the existing literature, factors of production in our model are flexible but are subject to diminishing returns. This allows us to highlight the tension between incentive provision and productive efficiency, which in turn generates systematic relationships between firm-level characteristics and workers' careers.

We also contribute to the vast literature on dynamic moral hazard; see Bolton and Dewatripont (2005, chapter 10) for a textbook treatment. As in efficiency-wage models (Shapiro and Stiglitz, 1984), our model assumes that wages are tied to jobs, giving rise to incentive rents. The efficiency-wage literature has studied the question of how firms can extract these incentive rents from workers by backloading pay within a given job (Lazear, 1979; Carmichael, 1985; Akerlof and Katz, 1989; Board, 2011; Lazear, Shaw, and Stanton, 2013; Fong and Li, 2014). In our model, backloading pay occurs across activity assignments, and our setting is indeed one in which the firm is able to extract all the surplus from workers. More importantly, we show that how a worker's pay is optimally backloaded (i.e., how his career progresses) is not determined in isolation. Rather, how a workers' pay is optimally backloaded depends on the firm's production technology and the careers of his coworkers. Our model therefore highlights how firm-level factors such as its production technology and its organizational demographics affect the dynamic moral-hazard problem at the worker level.

There is a sizeable literature looking at how the need to provide incentives interacts with organizational design (Williamson, 1967; Calvo and Wellisz, 1978; Qian, 1994; Mookherjee, 2013). In these models, workers remain in a fixed position within the firm, and the firm's monitoring technology is the key driver of its organizational structure. In our model, workers' positions within the hierarchy are not fixed, and their promotion opportunities determine their incentives. The need to provide incentives, therefore, affects the firm's optimal organizational structure.

Finally, there is a sociological literature on organizational demography that examines how the careers of individual workers progress within organizations and how the rate of career progression depends on the organization's span (Simon, 1951; White, 1970; Bartholomew, 1973; Keyfitz, 1973; Stewman and Konda, 1983; Rosenbaum, 1984; Stewman, 1986). Relative to this literature, our model emphasizes the active role a firm plays in shaping its workers' careers through both its personnel policies and organizational structure. Allowing for endogenous organizational responses can sometimes lead to different implications for workers' careers. For example, programs encouraging turnover of older workers increase the hiring of younger workers when the organizational structure is given, but it may have the opposite effect when the organizational structure is a choice.

2 The Model

A firm and a large mass of identical workers interact repeatedly. Time is discrete and denoted by t, and all players share discount factor $\delta \in (0, 1)$. We focus on the steady state and suppress time subscripts. Production requires two types of activities to be performed, and each worker can perform a single activity each period. A worker performing activity i in period t chooses an effort level $e_i \in \{0,1\}$ at cost $c_i e_i$. A worker who chooses $e_i = 0$ is said to *shirk*, and a worker who chooses $e_i = 1$ is said to *exert effort*. We refer to such a worker as *productive*. A worker's effort is his private information, but shirking in activity i is contemporaneously detected with probability q_i. If the firm employs masses N_1 and N_2 of productive workers in the two activities, revenues are $F(N_1, N_2)$. F is twice continuously differentiable, increasing, concave, and satisfies $F_{12} \geq 0$.

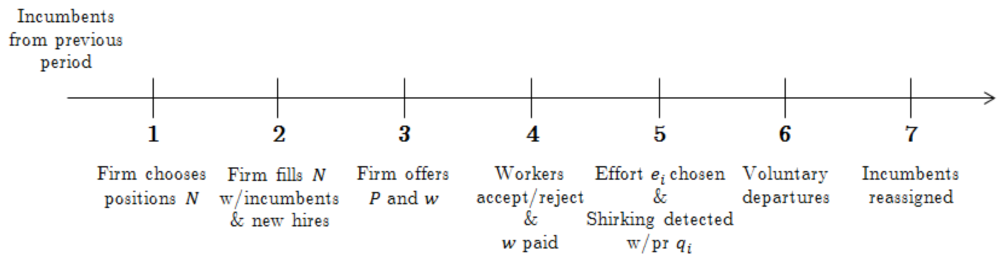

Figure 1: Timing of the stage game.

Figure 1 illustrates the timing of each period. The firm chooses the masses of positions N_1 and N_2 for each activity. The firm then fills these positions with incumbent workers and new hires, where we denote the mass of new hires for activity i as H_i, $i = 1, 2$. The firm offers each worker a contract (w_i, p_{ij}), $i, j = 1, 2$, that includes a *wage policy* and an *assignment policy* consisting of expected promotion, demotion, and retention patterns. We assume that wages are tied to activities, and denote the wage for activity i by w_i. The assignment policy is described by p_{ij}, which denotes the probability that a worker in activity i will take on activity j next period if he remains with the firm. We assume that a worker who is caught shirking is fired with probability 1, which constitutes an optimal penal code since it occurs only off the equilibrium path.

If a worker rejects the contract, he receives his outside option, yielding 0 utility. If he accepts the offer, the wage is paid and he chooses his effort level $e_i \in \{0,1\}$ at cost $c_i e_i$. If he chooses $e_i = 0$, he is caught shirking with probability q_i and fired. For workers not caught shirking, a fraction d_i of workers in activity i exogenously leave the firm. We refer to d_i as the *voluntary departure rate* of workers in activity i and assume that $d_1 + d_2 \leq 1$. Incumbent workers are reassigned according to the probabilities p_{ij}. If $p_{i1} + p_{i2} < 1$, some workers are asked to leave the firm and receive their outside utility. We refer to $1 - p_{i1} - p_{i2}$ as the *forced-turnover rate* for activity i.

3 Parallel-Careers Benchmark

To provide a benchmark against which to compare our results and to develop some useful notation and terminology, we begin by describing what we will refer to as the *parallel-careers benchmark.*

In this benchmark, the firm treats the two activities independently and offers a wage above the workers' outside options combined with the threat of termination following observed shirking in order to motivate effort. There is no worker mobility across activities.

Given a mass $\hat{N}_j$ of workers in activity j, the firm chooses N_i and w_i to solve the program:

$$\max_{N_i, w_i} F\left(N_i, \hat{N}_j\right) - w_i N_i,$$

subject to an individual-rationality constraint ensuring that the worker receives a greater payoff within the job than outside the job and an incentive-compatibility constraint ensuring that the worker prefers to choose $e_i = 1$ rather than $e_i = 0$. If the worker exerts effort in each period, he receives a total payoff of v_i in the job, where

$$v_i = w_i - c_i + (1 - d_i)\,\delta v_i.$$

That is, in each period, he receives the wage w_i and incurs the effort costs c_i. With probability d_i, he exogenously leaves the firm, but with the remaining probability, he remains in the job and receives v_i again the following period.

The worker will exert effort as long as

$$v_i \geq w_i + (1 - q_i)(1 - d_i)\,\delta v_i.$$

A worker who shirks avoids incurring the cost c_i but is caught and fired with probability q_i. A worker's motivation to work therefore derives from his expected future payoffs within the firm. Define the *incentive rents for activity* i as the minimum future payoffs necessary to satisfy the worker's incentive-compatibility constraint in activity i, and denote this value by R_i. The incentive-compatibility constraint can be rearranged to verify that

$$R_i = c_i / \left((1 - d_i)\,\delta q_i\right).$$

To maximize its profits, the firm chooses wages w_i, or equivalently, payoffs v_i, to ensure the incentive-compatibility constraint holds with equality. Given the resulting wage, the firm hires workers until the marginal revenue product of an additional worker is equal to this wage. Finally, the firm hires a mass of new workers into each activity to exactly offset the mass of workers who are exogenously separated from that activity. The resulting solution, which we refer to as the parallel-careers solution and denote with the superscript *pc*, is described in the following lemma.

LEMMA 0. *A firm maximizing its profits separately over the two activities chooses wages* $w_i^{pc} = c_i + (1 - (1 - d_i)\,\delta)\,R_i$ *to provide rents* $v_i^{pc} = R_i$ *to each worker performing activity* $i = 1, 2$*. The firm hires* $H_i^{pc} = (1 - d_i)\,N_i^{pc}$ *workers, where* $F_i\left(N_i^{pc}, N_j^{pc}\right) = w_i^{pc} > c_i$*.*

Lemma 0 is consistent with several observations of Shapiro and Stiglitz (1984). First, the firm has to pay wages exceeding workers' outside options to provide incentives. The resulting "efficiency wage" increases in the turnover rate d_i and decreases in the firm's monitoring ability, q_i. Second, the firm optimally chooses employment levels for each activity that are lower than the socially optimal level, which would satisfy $F_i = c_i$. The gap between the firm's employment-level choice and the socially optimal level is greater for activities that require higher incentive rents.

4 Managing Careers

In the parallel-careers benchmark, the firm chooses only a mass of workers to perform each activity and a wage paid to each of these workers. In this section, we study more general personnel policies that allow for reassignment across activities. We show that the firm always performs better by linking the activities together in the form of a career. We then characterize the firm's optimal choices and show that they lead to features characteristic of internal labor markets.

4.1 Preliminaries

The firm chooses wage, hiring, and assignment policies jointly to maximize its steady-state profits:

$$F(N_1, N_2) - w_1 N_1 - w_2 N_2.$$

As in the benchmark, denote v_i as the expected discounted payoff of a worker performing activity i. The firm maximizes its profits subject to the following constraints.

Promise-Keeping Constraints. Productive workers' payoffs have to be equal to the sum of their current payoffs and their continuation payoffs:

$$v_1 = w_1 - c_1 + (1 - d_1)\,\delta\,(p_{11} v_1 + p_{12} v_2); \tag{PK-1}$$

$$v_2 = w_2 - c_2 + (1 - d_2)\,\delta\,(p_{21} v_1 + p_{22} v_2). \tag{PK-2}$$

Individual-Rationality Constraints. Workers prefer not to take their outside options if

$$v_1 \geq 0; \tag{IR-1}$$

$$v_2 \geq 0. \tag{IR-2}$$

Incentive-Compatibility Constraints. Workers prefer to exert effort if the following conditions hold:

$$w_1 - c_1 + (1 - d_1)\,\delta\,(p_{11} v_1 + p_{12} v_2) \geq w_1 + (1 - q_1)(1 - d_1)\,\delta\,(p_{11} v_1 + p_{12} v_2);$$

$$w_2 - c_2 + (1 - d_2)\,\delta\,(p_{21} v_1 + p_{22} v_2) \geq w_2 + (1 - q_2)(1 - d_2)\,\delta\,(p_{21} v_1 + p_{22} v_2),$$

where we use the fact that if the worker leaves the firm, he receives a payoff of 0. Equivalently, future payoffs have to exceed activity i's incentive rents:

$$p_{11}v_1 + p_{12}v_2 \geq c_1/\left(\left(1-d_1\right)\delta q_1\right) = R_1; \tag{IC-1}$$

$$p_{21}v_1 + p_{22}v_2 \geq c_2/\left(\left(1-d_2\right)\delta q_2\right) = R_2, \tag{IC-2}$$

where R_i is the incentive rent for activity i. Without loss of generality, we assume that $R_2 \geq R_1$.

Flow Constraints. In the steady state, the number of workers in a particular activity must remain constant. Given the hiring and assignment policies, the following constraints ensure that the mass of workers flowing into each activity equals the mass of positions for that activity:

$$\left(1-d_1\right)p_{11}N_1 + \left(1-d_2\right)p_{21}N_2 + H_1 = N_1; \tag{FL-1}$$

$$\left(1-d_1\right)p_{12}N_1 + \left(1-d_2\right)p_{22}N_2 + H_2 = N_2, \tag{FL-2}$$

where $H_i \geq 0$ is the mass of new workers hired into activity i. In addition, since p_{ij} are probabilities, they must be nonnegative, and

$$p_{i1} + p_{i2} \leq 1, \text{ for } i = 1, 2.$$

A fraction of workers who are neither caught shirking nor exogenously separated from the firm are fired if $p_{i1} + p_{i2} < 1$.

We solve the firm's problem in two steps. First, we fix the number of positions for each activity, and we solve for the firm's cost-minimizing levels of p_{ij}, H_i, and v_i. We then allow the firm to optimize over N_1 and N_2. We refer to the ratio N_1/N_2 as the firm's *span* and the pair (N_1, N_2) as the firm's *organizational structure*. The vector $\mathbf{H} = [H_i]_i$ is the firm's *hiring policy*, and the rent vector $v = [v_i]_i$ determines the firm's *wage policy* $\mathbf{w} = [w_i]_i$ for a given *assignment policy* $\mathbf{P} = [p_{ij}]_{ij}$. The values $1 - p_{i1} - p_{i2}$ represent the forced turnover rate in activity i to leave the firm, so the assignment policy $\mathbf{P}$ represents the firm's *promotion*, *demotion*, and *retention policies*. If $p_{i1} + p_{i2} = 1$, we say that activity i has *full job security*; that is, a worker performing activity i departs the firm only for exogenous reasons, unless he is caught shirking. We refer to a collection $(\mathbf{H}, \mathbf{w}, \mathbf{P})$ as a *personnel policy*.

4.2 Optimal Personnel Policy

In this section, we characterize the optimal personnel policy given an organizational structure (N_1, N_2). Given (N_1, N_2), the firm chooses a personnel policy $(\mathbf{H}, \mathbf{w}, \mathbf{P})$ to solve the program:

$$W\left(N_1, N_2\right) = \min_{(\mathbf{H}, \mathbf{w}, \mathbf{P})} w_1 N_1 + w_2 N_2,$$

subject to $(PK - i)$, $(IR - i)$, $(IC - i)$, and $(FL - i)$. That is, the firm chooses hiring, wage, and assignment policies to minimize the steady-state wage bill. Throughout this section, we will focus on organizational structures in which there are more workers performing the low-rent activity (i.e., $N_1 \geq N_2$). We characterize the full solution in the appendix and discuss at the end of this section the differences that arise when $N_2 > N_1$. In the next section, we solve for the optimal organizational structure, and we specify conditions on the production function so that the optimal organizational structure indeed has $N_1 \geq N_2$.

Throughout this section, we will assume (and formally verify in the appendix) that under the optimal personnel policy, whenever $N_1 \geq N_2$, the rents provided in activity 2 exceed those provided in activity 1 (i.e., $v_2^* > v_1^*$). For reasons that will soon become clear, we refer to activity 1 as the *bottom job* and activity 2 as the *top job*. We also refer to workers who perform activity 1 as *bottom workers* and those who perform activity 2 as *top workers*. Since $d_1 + d_2 \leq 1$ and $N_1 \geq N_2$, we necessarily have that $N_2 d_2 \leq N_1 (1 - d_1)$, so that there are enough incumbent bottom workers to fill all top-job vacancies generated by voluntary turnover. Our first result shows that the firm will never hire directly into the top job.

LEMMA 1. *All new workers are hired into the bottom job (i.e.,* $H_2^* = 0$*).*

To see why firms prefer to hire workers into the bottom job, notice that a vacancy in the top job can be filled either by directly hiring into the top job or by hiring into the bottom job and promoting an incumbent bottom worker. Hiring directly into the top job requires the firm to provide a rent of v_2^* to the new worker. In contrast, hiring into the bottom job and promoting an incumbent bottom worker only requires the firm to provide a rent of v_1^* to the new worker. Both policies preserve the flow constraint, since the vacancy in the top job is filled and the mass of bottom workers remains constant. Hiring into the bottom job also makes the incentive-compatibility and participation constraints for bottom workers easier to satisfy, because it involves a higher promotion probability. Promoting from within helps motivate bottom workers using the rents associated with the top job, which in turn allows the firm to lower the wages associated with the bottom job.

Next, we describe workers' careers within the firm. There will be two important cases to consider, which are related to the rents that are freed up by voluntary departures at the top. Consider the parallel-careers benchmark in which there are no promotions, and each activity is associated with full job security and is paid a wage that corresponds to its incentive rents. At the end of any period, a mass $d_2 N_2$ of workers depart from the top, which frees up an amount $d_2 N_2 R_2$ of rents that may be reallocated. Additionally, at the end of the period, there is a mass $(1 - d_1) N_1$ of incumbent bottom workers who must be promised rents R_1 to exert effort. We say that there are *sufficient separation rents* if $d_2 N_2 R_2 \geq (1 - d_1) N_1 R_1$. In this case, the prospect of receiving rents from exogenous turnover of the top job is sufficient to motivate the workers at the bottom job. If this condition is not satisfied, we say that there are *insufficient separation rents.* The next lemma describes workers' careers when there are sufficient separation rents.

LEMMA 2. *When there are sufficient separation rents, in an optimal personnel policy, bottom workers receive zero rents, and top workers receive the incentive rents associated with activity* 2*. There are no demotions, and workers receive full job security.*

Lemma 2 illustrates the benefits of using promotions to reduce rents given to new workers. In the parallel-careers benchmark, high wages motivate workers and also determine their equilibrium payoffs. By using promotions, the firm can separate incentive provision from equilibrium payoffs for bottom workers. Since top workers are never promoted, they must receive at least the incentive rents for activity 2 in order to exert effort. When there are sufficient separation rents, promotion prospects alone provide enough motivation for bottom workers, so that their incentive-compatibility constraints are slack. The firm then sets the bottom wage just high enough to induce participation, leaving bottom workers with no rents. Bottom workers' per-period payoffs are lower than their outside options, but they are willing to work for the firm, because of the prospect of being promoted to the top job.

If top workers were demoted or forced to leave the firm with positive probability, the incentive rents for activity 2 would not be sufficient to motivate them. Since they receive the incentive rents for activity 2 under the optimal personnel policy, it must therefore be the case that they are never demoted, and they receive full job security. For bottom workers, full job security is optimal, but not uniquely so. As long as the promotion probability of bottom workers at the *beginning* of each period remains unchanged, workers are motivated, and the firm's wage bill is the same. Any hiring or firing costs would make full job security for bottom workers uniquely optimal. This is because full job security for bottom workers minimizes the mass of workers who are hired and fired.

Workers' career patterns are different in firms in which there are insufficient separation rents. We explore these patterns in the next lemma.

LEMMA 3. *When there are insufficient separation rents, in an optimal personnel policy, bottom workers receive zero rents, and top workers receive rents in excess of the incentive rents for activity* 2*. There are no demotions, bottom workers receive full job security, and there is forced turnover at the top.*

When there are insufficient separation rents, the personnel policies described in Lemma 2 no longer provide enough motivation for bottom workers. To increase the incentives for bottom workers, the firm could in principle pay higher wages at the bottom. Lemma 3 shows that doing so is never optimal—in the optimal personnel policy, bottom workers receive zero rents. The firm provides additional motivation entirely by increasing bottom workers' promotion prospects. To do this, the firm fires top workers with positive probability in each period and offers them rents that exceed the incentive rents for activity 2. This increase in turnover at the top allows the firm to increase the promotion prospects for bottom workers. Coupled with the associated increase in rents upon promotion, such a policy maintains motivation for both top workers and bottom workers.

To see in another way why the firm prefers to use promotion incentives rather than efficiency wages to motivate bottom workers, notice that if higher wages are paid at the bottom, the firm must be giving rents to new workers. Doing so constitutes a pure loss for the firm. In contrast, the firm can recapture increased wages for top workers by lowering wages for bottom workers. Raising wages for top workers backloads a worker's pay and therefore is more effective than offering high wages throughout the firm. Moreover, if the firm offers rents that exceed the incentive rents for activity 2 for the top job, top workers' incentive constraints would be slack if they were given full job security. The firm can therefore reduce top workers' job security as well as increase bottom workers' promotion prospects.

Further, forced-turnover policies weakly dominate demotions for top workers. Both policies create promotion opportunities for bottom workers, but they also reduce the value that workers place on the top job. The relative amount by which they do so depends on how top workers' outside options compare to the value of the bottom job, which under the optimal personnel policy is equal to the bottom workers' outside options. Forced turnover is therefore preferred whenever top workers' outside options exceed bottom workers' outside options. For demotions to be optimal, it has to be the case that bottom workers' outside options are greater than top workers' outside options. In our model, both are zero.

Finally, it is worth remarking that optimal wages, promotion prospects, and forced-turnover rates depend on (N_1, N_2) only through the span N_1/N_2. This is because for any (N_1, N_2), hiring only occurs at the bottom, and bottom workers receive zero rents. Wages at the bottom are therefore determined by bottom workers' promotion prospects, which depend on the firm's span. Wages at the top are determined by the incentive rents for activity 2 and the forced-turnover rate, which also depend on the firm's span.

Proposition 1 summarizes the main features of an optimal personnel policy.

PROPOSITION 1. *An optimal personnel policy has the following features: (i) Hiring occurs only* *into the bottom job. (ii) There is a well-defined career path: Bottom workers stay at the bottom job or are promoted. Top workers are never demoted but may be fired. (iii) Bottom-job wages correspond to rents that are lower than the incentive rents for activity* 1. *Top-job wages correspond to rents that exceed the incentive rents for activity* 2 *whenever there are insufficient separation rents. (iv) Wages, promotion rates, and forced-turnover rates depend on* (N_1, N_2) *only through the span* N_1/N_2.

Proposition 1 characterizes optimal personnel policies, given the firm's organizational structure and therefore results in a **labor-cost function**, $W(N_1, N_2)$. We now discuss several properties of the labor-cost function. Given (N_1, N_2), the expressions for optimal wages and for the labor-cost function depend on whether there are sufficient separation rents (i.e., whether $d_2 N_2 R_2 \geq (1 - d_1) N_1 R_1$). There are sufficient separation rents if $N_1/N_2 \leq \kappa$, where

$$\kappa \equiv (R_2/R_1) \cdot (d_2/(1 - d_1)).$$

That is, there are sufficient separation rents whenever the firm's span is low and/or the voluntary turnover rate of the top job is high. These expressions for wages and for the labor-cost function are described in the following corollary to Proposition 1.

COROLLARY 1. *The following is true when* $N_1 \geq N_2$*:*

(*i*) *When there are sufficient separation rents, bottom wages are* $w_1^* = c_1 - \delta(1 - d_1) R_1 \kappa N_2/N_1$ *and top wages are* $w_2^* = c_2 + (1 - \delta)(1 - d_2) R_2 + d_2 R_2$. *The labor-cost function is*

$$W(N_1, N_2) = c_1 N_1 + c_2 N_2 + (1 - \delta) N_2 R_2.$$

(*ii*) *When there are insufficient separation rents, bottom wages are* $w_1^* = c_1 - \delta(1 - d_1) R_1$ *and top wages are* $w_2^* = c_2 + (1 - \delta)(1 - d_2) R_2 + d_2 (N_1/N_2) \kappa^{-1} R_2$. *The labor-cost function is*

$$W(N_1, N_2) = c_1 N_1 + c_2 N_2 + (1 - \delta)((1 - d_1) N_1 R_1 + (1 - d_2) N_2 R_2).$$

In each of the regions described in Corollary 1, the labor-cost function is linear in N_1 and N_2, so the coefficient on N_i has a natural interpretation as the marginal cost to the firm of adding a position in activity i. Under the Neoclassical model of labor supply in which there are no incentive problems, this coefficient would be equal to the wage for workers in activity i, which would be equal to the associated effort cost c_i, since workers' outside options are 0 and they are on the long side of the market.

In contrast, when effort is not contractible, the marginal cost accounts for the effect that adding an additional position for activity i affects the firm's optimal personnel-policy problem, which in turn depends on whether there are sufficient separation rents. When there are sufficient separation rents, the marginal cost of adding a position for activity 1 coincides with the Neoclassical costs of adding the position, c_1, which in turn exceeds the bottom wage, because compensation is backloaded in workers' careers. When there are insufficient separation rents, adding another position at the bottom reduces the promotion prospects for bottom workers and therefore requires that the firm adjust its personnel policies in order to keep bottom workers motivated. The resulting effective marginal cost of adding such a position is then greater than c_1. Relatedly, there are benefits of adding positions at the top that exceed the marginal revenue product of such positions, since additional positions at the top create promotion opportunities for workers at the bottom, in turn relaxing the firm's optimal personnel-policy problem.

Finally, we briefly remark on optimal personnel policies when there are more positions for workers performing the high-rent activity than there are for the low-rent activity (i.e., when $N_2 > N_1$). Under the optimal personnel policy, activity 1 becomes the top job, and all hiring occurs into the bottom job in which activity 2 is performed. Further, there are no demotions, and $v_2^* = 0$. In contrast to the case when $N_1 > N_2$, the firm always puts in place forced-turnover policies. Moreover, the incentive constraint may be slack for workers at the top—in this case, it is optimal to promote all incumbent bottom workers in every period, and the firm will push out just enough top workers to make this feasible.

If we define the cutoff

$$\sigma \equiv (1 - d_2) R_2 / (R_2 - (1 - d_1) R_1),$$

then when $N_1/N_2 > \sigma$, the labor-cost function is

$$W(N_1, N_2) = N_1 c_1 + N_2 c_2 + (1 - \delta) N_1 R_1,$$

and when $N_1/N_2 \leq \sigma$, the labor-cost function is

$$W(N_1, N_2) = c_1 N_1 + c_2 N_2 + (1 - \delta)((1 - d_1) N_1 R_1 + (1 - d_2) N_2 R_2).$$

In Section 5, we impose conditions on the production function so that the firm will always choose to have more workers performing the low-rent activity (i.e., so that $N_1 \geq N_2$).

5 Optimal Production

Given the labor-cost function and the production function, we use standard tools from Neoclassical production theory to characterize the *optimal organizational structure* (N_1^*, N_2^*). We first characterize a few general properties of the optimal organizational structure. We then explore two specific topics that will be relevant for our empirical discussion in Section 6. First, we examine how changes in the voluntary turnover rate affect the firm's optimal span and production level. We then draw connections between the firm's optimal production level and its optimal span.

We write the firm's revenues as the product of its output price and its output: $F(N_1, N_2) = P \cdot f(N_1, N_2)$, where we assume f is increasing, concave, and satisfies $f_{12} \geq 0$. Given the firm's labor-cost function $W(N_1, N_2)$, the firm solves

$$\max_{N_1, N_2} P \cdot f(N_1, N_2) - W(N_1, N_2).$$

We analyze this problem in two steps. First, the firm chooses the cost-minimizing organizational structure necessary to produce output y. The cost-minimizing organizational structure in turn determines the optimal personnel policy. Next, the firm chooses the *optimal production level* y^*.

Given output level y, the firm chooses $(N_1^*(y), N_2^*(y))$ to minimize costs:

$$C(y) = \min_{N_1, N_2} W(N_1, N_2),$$

subject to $f(N_1, N_2) \geq y$. We now make an assumption on f to ensure optimal production involves more workers performing the low-rent activity than the high-rent activity. Define the marginal rate of technical substitution as $MRTS(N_1, N_2) = f_1/f_2$. Our assumption ensures that at any point $N_1 \leq N_2$, the firm can always lower its costs by decreasing N_2 and increasing N_1.

ASSUMPTION 1 (Production Favors Activity 1). *For any* N, $MRTS(N, N) > (c_1 + R_2)/c_2$.

Assumption 1 is satisfied by, for instance, a Cobb-Douglas production function $f(N_1, N_2) = N_1^{\alpha_1} N_2^{\alpha_2}$ with $(\alpha_1/\alpha_2) > (c_1 + R_2)/c_2$. From Corollary 1, we know $W(N_1, N_2)$ is piecewise linear in (N_1, N_2), and the coefficients on N_1 and N_2 depend on the region in which the firm operates. Figure 2 below depicts the producer-theory approach to the firm's cost-minimization problem. Isocost curves are piecewise-linear with different coefficients on either side of the three boundaries.

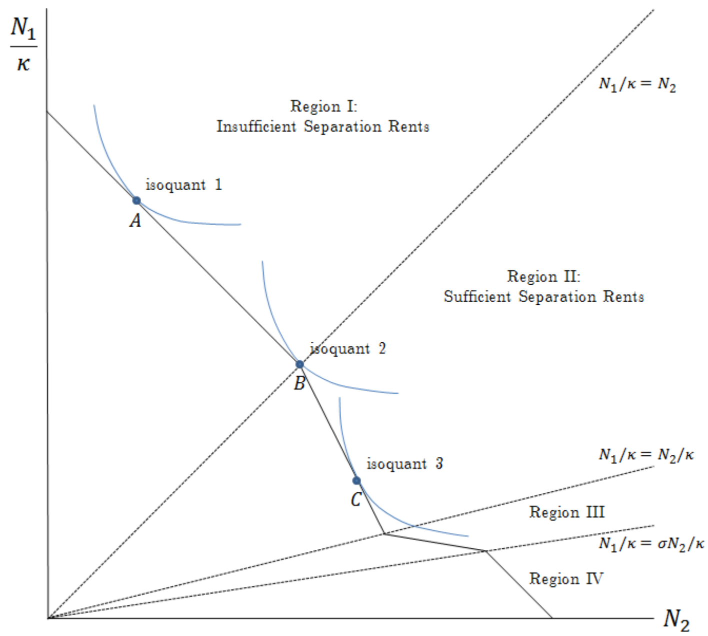

Figure 2: Producer-theory approach to a firm's cost-minimization problem. The isocost curve is piecewise linear with different coefficients on either side of the dotted boundaries. The isoquants represent different production technologies, with higher-numbered isoquants representing production technologies increasingly favoring activity 2 relative to activity 1. Notice that the vertical axis on the graph has been rescaled to emphasize the relevant regions.

The first aspect of optimal production to notice is that when production favors activity 1, firms will never produce in Region III or Region IV. Each of the three isoquants depicted in Figure 2 represents a different production technology producing the same level of output y. Isoquant 1 is an activity-1-heavy production technology and will favor production at point A at which the firm operates with a large span and has insufficient separation rents. Isoquant 3 is an activity-2-heavy production technology and will favor production at point C at which the firm operates with a small span and has sufficient separation rents. At each of the non-boundary points, the marginal rate of technical substitution is equal to the cost ratio, W_1/W_2. Corollary 2 below, which follows directly from Corollary 1, formally describes the marginal rate of technical substitution for the three cases.

COROLLARY 2. *Given a production level y, the optimal organizational structure satisfies $N_1^*(y) > N_2^*(y)$, and the following conditions hold:*

(*i*) *When $N_1^*(y) < \kappa N_2^*(y)$, the marginal rate of technical substitution is given by*

$$MRTS\left(N_1^*, N_2^*\right) = \frac{\delta - \delta d_2}{1 - \delta d_2}\frac{c_1}{c_2}.$$

(*ii*) *When $N_1^*(y) = \kappa N_2^*(y)$, the marginal rate of technical substitution satisfies*

$$MRTS\left(N_1^*, N_2^*\right) \in \left[\frac{\delta - \delta d_2}{1 - \delta d_2}\frac{c_1}{c_2}, \frac{c_1}{c_2}\right].$$

(iii) When $N_1^*(y) > \kappa N_2^*(y)$, the marginal rate of technical substitution is given by

$$MRTS(N_1^*, N_2^*) = \frac{c_1}{c_2}.$$

Corollary 2 shows that firms producing in the insufficient-separation-rents region choose an organizational structure for which the marginal rate of technical substitution is high—when bottom workers' promotion opportunities are limited, the firm optimally creates relatively more positions at the top. When there are insufficient separation rents, creating additional positions at the top generates a positive spillover effect, because each additional position improves the promotion opportunities of bottom workers and, as a result, allows the firm to lower top wages. When there are sufficient separation rents, in contrast, top wages are independent of the number of top positions, and the benefits to creating additional top positions is therefore smaller.

Corollary 2 determines the optimal organizational structure $(N_1^*(y), N_2^*(y))$ and minimized production cost $C(y)$ for each production level y. Varying y traces out a production-expansion path. The firm maximizes its profits by choosing y to solve the following unconstrained program:

$$\max_y Py - C(y).$$

The optimal production level solves $C'(y^*) = P$. Given the optimal production level y^*, the optimal organizational structure is $(N_1^*(y^*), N_2^*(y^*))$, and the associated wage levels are w_1^* and w_2^*.

If effort were directly contractible, optimal production would equate the marginal revenue product of each position to the wages associated with that position (w_i^*). Denote the marginal revenue product at the optimum by $MRP_i^* = P \cdot f_i(N_1^*(y^*), N_2^*(y^*))$. In contrast, when effort is not contractible, top wages exceed the marginal revenue product for top positions and bottom wages lie below the marginal revenue product for bottom positions.

COROLLARY 3. *At the optimum, $w_1^* < MRP_1^*$ and $w_2^* > MRP_2^*$.*

The departure from standard Neoclassical optimality conditions arises, because wages are backloaded across activities in order to motivate workers. This backloading implies that creating additional top positions relaxes the firm's incentive problem, and therefore the firm will go beyond $w_2^* = MRP_2^*$ and create additional positions. Similarly, creating additional bottom positions tightens the firm's incentive problem, so the firm will stop shy of $w_1^* = MRP_1^*$.

In the next two subsections, we explore how optimal production provides connections between environmental characteristics such as the voluntary departure rate or product demand and the firm's optimal organization and personnel policies.

5.1 The Role of Turnover

In contrast to standard Neoclassical production theory, worker turnover plays an important role in shaping optimal production precisely because it affects trade-offs in the firm's incentive problem. In this subsection, we explore the role that the voluntary turnover rate at the top, d_2, plays in determining optimal production.

We first note that the voluntary turnover rate at the bottom, d_1, has no effect on production. An increase in d_1, on the one hand, makes incentive provision more difficult for bottom workers, since it lowers their effective discount factors. On the other hand, it implies that promotion prospects are higher for bottom workers, since fewer bottom workers remain at the end of each period. These two effects exactly cancel each other out in our model.

Given a turnover rate d_2^* and holding total labor costs fixed, Figure 3 below depicts the effects of an increase in the voluntary turnover rate at the top. Define $\kappa^* = (c_2/c_1) \cdot (d_2^*/(1-d_2^*))$. Increasing d_2 away from d_2^* increases the threshold span κ below which the firm operates in the sufficient-separation-rents region. Moreover, in the sufficient-separation-rents region, an increase in d_2 increases the wage necessary to motivate top workers, with no offsetting effect on the motivation of bottom workers, leading to a counterclockwise rotation of the isocost line.

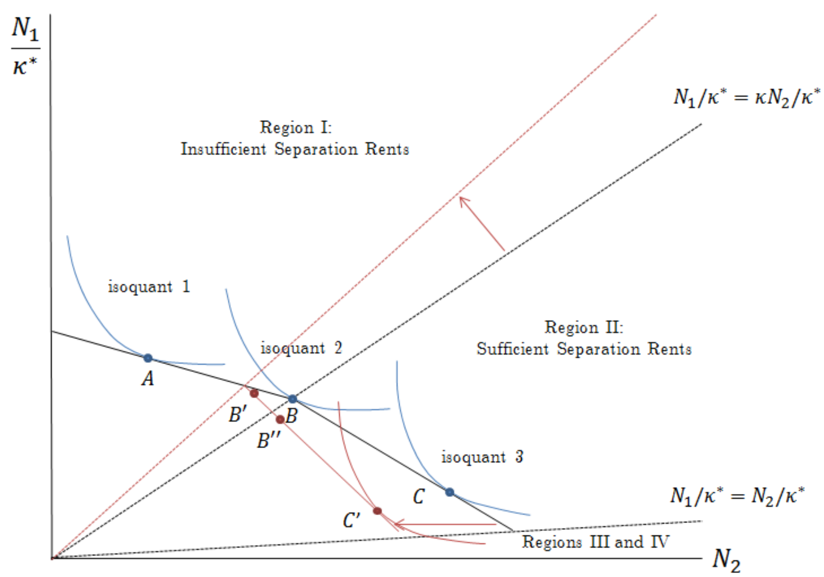

Figure 3: This figure examines the effects of an increase in the voluntary departure rate at the top. Holding labor costs constant, this increase rotates the boundaries between the ISR and SSR regions counterclockwise, and it rotates the isocost curve in the SSR region clockwise. If point A was optimal, then it will remain optimal. If point B was optimal, then the new optimum will be either B' or B". If point C was optimal, then the new optimum will be C'.

Firms operating at points like A are unaffected by an increase in the voluntary-turnover rate at the top, since they optimally offset this increase by decreasing the forced-turnover rate. Firms operating at points like C optimally reduce the number of positions at the top and the bottom and instead produce at a point like C'. Firms operating at boundary span points like B will reduce production and will reduce N_2, but depending on the production technology, they might increase or decrease N_1. These effects are summarized in Corollary 4 below.

COROLLARY 4. The following is true: (i) if $N_1^*/N_2^* > \kappa$, then $dN_1^*/dd_2 = dN_2^*/dd_2 = 0$; (ii) if $N_1^*/N_2^* = \kappa$, then $dN_2^*/dd_2 < 0$; and (iii) if $N_1^*/N_2^* < \kappa$, then $dN_1^*/dd_2 \leq 0$ and $dN_2^*/dd_2 < 0$.

Part (i) shows that an increase in the voluntary turnover rate d_2 has no effect on the organizational structure when there are insufficient separation rents. To see this, notice that for any organizational structure (N_1, N_2), there are insufficient separation rents whenever the voluntary turnover rate is below a threshold level $\bar{d}_2$. Under the optimal personnel policy, $\bar{d}_2$ becomes the lower bound for the firm's total (voluntary plus forced) turnover rate at the top. That is, the firm puts in place forced turnover policies if $d_2 < \bar{d}_2$, so that the total turnover rate at the top remains at $\bar{d}_2$. The voluntary turnover rate, therefore, does not affect the organizational structure or workers' careers in this case.

In Parts (ii) and (iii), the firm does not force turnover, and changes in d_2 affect the firm's organizational structure. In particular, a higher d_2 decreases the effective discount factor of workers at the top. The firm must therefore increase wages at the top and, in turn, reduce the number of positions at the top (i.e., $dN_2^*/dd_2 < 0$). More importantly, a higher turnover rate at the top causes the firm to adjust the number of positions at the bottom. This spillover effect operates through both a production channel and an incentive channel. In Part (iii), bottom workers' incentive constraints are slack, so the spillover operates solely through the production channel: complementarity in production implies that whenever there are fewer top positions, the firm also reduces the number of bottom positions (i.e., $dN_1^*/dd_2 \leq 0$). In Part (ii), however, when the bottom worker's incentive constraints are binding, the incentive channel is also in force. A higher d_2 relaxes bottom worker's incentive constraints, because it allows the firm to improve bottom workers' promotion prospects given N_1 and N_2. The firm can therefore create additional bottom positions without violating existing workers' incentive constraints. This additional effect can dominate, and N_2^* can potentially increase in d_2.

5.2 Production Level and Organizational Structure

We now discuss how the production level (y) affects the firm's optimal span $\left(N_1^*(y)/N_2^*(y)\right)$. Since the firm's optimal personnel policies are determined solely by the firm's span, this allows us to study how the production level affects which personnel policies the firm puts in place.

There is broad evidence that larger firms tend to have larger spans (see Rushing (1966), Blau and Schoenherr (1971), Kasarda (1974), and Colombo and Delmastro (1999) for cross-sectional evidence; see Caliendo, Monte, and Rossi-Hansberg (Forthcoming) for more recent evidence that, conditional on organizational structure, firms expand by increasing their span.) We will therefore focus on nonhomothetic production technologies for which *expansion favors activity 1.* Assumption 2 provides sufficient conditions for a production function f to have an increasing and convex production-expansion path. Extreme production functions in which there is a fixed number of top positions in the firm, such as those in Zabojnik and Bernhardt (2001) and DeVaro and Morita (2013), can be viewed as a limiting case of Assumption 2.

ASSUMPTION 2 (Production Expansion Favors Activity 1). For all $k \geq 0$ and for some $\varepsilon > 0$,

$$kN_2 \left.\frac{\partial MRTS}{\partial N_1}\right|_{N_1=kN_2} + N_2 \left.\frac{\partial MRTS}{\partial N_2}\right|_{N_1=kN_2} \geq \varepsilon.$$

Assumption 2 ensures that the marginal rate of technical substitution between N_1 and N_2 strictly increases along any ray from the origin. In other words, as production expands, N_2 becomes a worse substitute for N_1 in production. Assumptions 1 and 2 are both satisfied by, for example, a Stone-Geary production function $f(N_1, N_2) = N_1^{\alpha_1}(N_2 - \gamma)^{\alpha_2}$ with $(\alpha_1/\alpha_2) > (c_1 + R_2)/c_2$, $\gamma > 0$, and a minimum production requirement arising from fixed costs of production.

When production expansion favors activity 1, the associated production-expansion path will be convex within each of the two regions, and it will be linear on the boundary. Figure 4 below depicts a production-expansion path for such a production function.

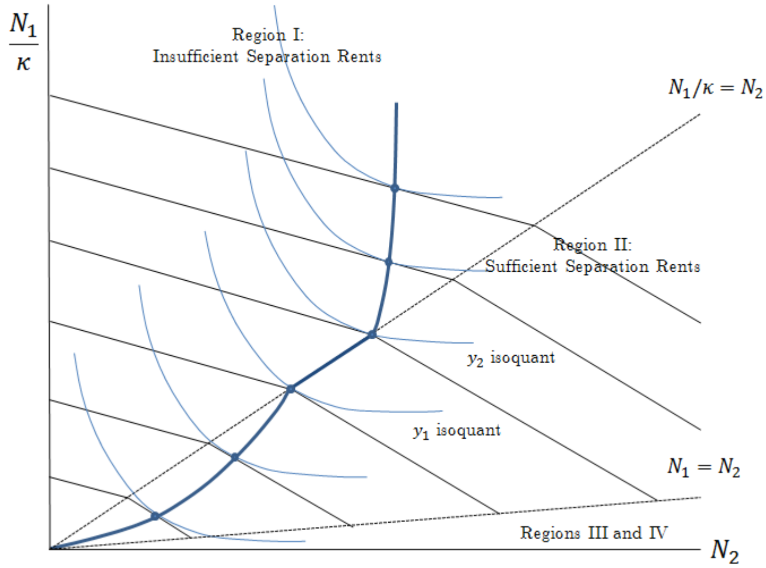

Figure 4: This figure plots a production-expansion path for a nonhomothetic production technology for which expansion favors activity 1. Firms with high production levels optimally operate in the insufficient-separation-rents region. Firms with low production levels optimally operate in the sufficient-separation-rents region. Firms with intermediate production levels optimally operate on the boundary between the two regions.

Figure 4 highlights the result, summarized in Corollary 5, that the firm's optimal production level determines whether it produces in the sufficient-separation-rents region or the insufficient-separation-rents region. As a result, the firm's production level determines the firm's span and therefore its optimal personnel policies.

COROLLARY 5. *Suppose production expansion favors activity 1. Then there exist two cutoffs,* y_1 *and* y_2*, such that the following is true:* (*i*) *if* $y^* < y_1$*, the firm's optimal span is* $N_1^*/N_2^* < \kappa$*;* (*ii*) *if* $y^* \in [y_1, y_2]$*, the firm's optimal span is* $N_1^*/N_2^* = \kappa$*; and* (*iii*) *if* $y^* > y_2$*, the firm's optimal span is* $N_1^*/N_2^* > \kappa$*.*

We can compare the optimal personnel policies of firms that have high production levels and those that have low production levels. Suppose there is a small firm (denoted by superscript S) that operates at $y^{*S} < y_1$ and a large firm (denoted by superscript L) that operates at $y^{*L} > y_2$. Corollary 6 summarizes key differences in the optimal personnel policies for these two firms.

COROLLARY 6. *Suppose production expansion favors activity 1. Wages are higher for both positions at large firms relative to small firms. Promotion probabilities are higher at small firms, and large firms put in place forced-turnover policies.*

In the next section, we link the predictions of Corollary 6 to the empirical literature on how the careers of workers differ between small and large firms.

6 Discussion of Empirical Implications

Our model delivers a rich set of predictions that are consistent with a broad pattern of evidence, and it provides a framework to think about how the need to manage workers' promotion prospects interacts with firm size and the firm's demographics. In Section 6.1, we highlight a number of predictions that accord with a host of stylized facts regarding internal labor markets. These predictions are core to the individual dynamic moral hazard problems between the firm and each worker.

In Sections 6.2 and 6.3, we then highlight two sets of predictions that go beyond the standard facts and arise precisely because workers' careers are managed at the firm level and therefore involve trade-offs among workers. The first set of implications relates the demographics of the workers to the organization of the firm. The second set of implications examines how firm size affects a firm's personnel policies and, relatedly, the career dynamics of its workers. Each of the facts we discuss is likely due to and consistent with many factors beyond the scope of our model, but taken as a whole, they are supportive of our model's main mechanisms and trade-offs.

6.1 Features of Internal Labor Markets

Proposition 1 shows that optimal personnel policies give rise to hiring and job-assignment patterns that are consistent with the functioning of internal labor markets.

OBSERVATION 1. *Employees perform different activities at different stages of their career with the firm.*

A core force in our model is the optimal backloading of incentive rents through activity assignments. This force offers a clear rationale for why organizations offer careers internally, an idea that was advanced in Doeringer and Piore's (1971) seminal work on internal labor markets. Of course, job ladders within firms can–and clearly do–arise for many reasons, including firm-specific human capital acquisition and employer learning about worker skills. We highlight here that another natural reason they might arise is because they serve as an efficient way of providing incentives. To the extent that job ladders arise as a way of providing incentives, they will necessarily have a number of additional features, which we will describe below.

OBSERVATION 2. *Employees begin their careers in the bottom job.*

According to Doeringer and Piore (1971), "Entry into [internal labor] markets is limited to particular jobs or ports of entry." Indeed, ports of entry appear to be the rule in some industries and countries. For example, Milgrom and Roberts (1992) report that "airlines used a strict system of hiring only at the bottom of the job ladder." As a result, even experienced pilots who lost their jobs due to the industry shakeout in the late 1980s and early 1990s started over at the bottom when they changed airlines. Imai and Hiroyuki (1988) observe that entry into Japanese firms is limited in the sense that most hiring is done at the time of graduation and mid-career ports of entry are almost nonexistent. A port of entry is an extreme feature of a personnel policy that is, of course, not present in every firm. For example, Baker, Gibbs, and Holmstrom (1994a) study detailed personnel records from a large U.S. firm and find that 25% of workers filling higher positions in the firm are hired externally. We discuss below how our model's backloading mechanism generates a port of entry as a stark limiting case of insider bias in hiring.

OBSERVATION 3. *Workers are never demoted.*

With a couple of recent exceptions (Dohmen et al., 2004; Lin, 2005), studies of detailed personnel data suggest that demotions are rare. For example, in Baker, Gibbs, and Holmstrom's (1994a) study, demotions almost never occur. Seltzer and Merrett (2000) report similar findings using data from an Australian bank, and Treble et al. (2001) find that demotions are rare in a British service-sector firm. In our model, a firm might in principle want to demote top workers in order to create opportunities for bottom workers. However, because the outside option of top workers is weakly higher than the payoffs bottom workers receive, top workers are better off being forced out of the firm than being demoted. In turn, the firm is better off adopting forced-turnover policies rather than demoting top workers.

OBSERVATION 4. *There are wage jumps at promotion.*

Many studies have found that promotions are associated with large wage increases (Murphy, 1985; Lazear, 1992; Baker, Gibbs, Holmstrom 1994a,b; McCue, 1996). The wage increases may result from a number of factors such as human capital accumulation (Demougin and Siow, 1994; Gibbons and Waldman, 1999a) and signaling of ability to the outside market (Waldman, 1984; Bernhardt, 1995). Lazear and Rosen's (1981) labor-market tournament model provides one of the first incentive-based explanations for large wage increases upon promotion—these wage increases are used to provide incentives for effort for workers at the bottom of the job ladder. But, in principle, the wage increases could have taken the form of a large one-time bonus. In our model, these wage increases optimally serve not only to provide incentives for bottom workers, but also for top workers. In other words, if a worker is willing to work hard to get a promotion, he will also be willing to work hard to keep the job to which he has been promoted.

OBSERVATION 5. *Wages at the bottom are below workers' marginal revenue product. Wages at the top exceed workers' marginal revenue product.*

Corollary 3 shows that at the conditionally optimal scale, bottom workers' wages are below their marginal revenue product. This reluctance to hire workers whose marginal revenue product exceeds their wages arises because hiring an additional worker at the bottom reduces the promotion prospects of *other* workers at the bottom, giving rise to a shadow cost of additional bottom positions. This wedge between marginal revenue product and wages can manifest itself as often-lamented "headcount restriction" policies that human-resource departments put in place.

On the flip side, top workers' wages exceed their marginal revenue product. Taken together with the prediction that workers start with the bottom job and receive a wage below their marginal product of labor, this implies that wage growth increases faster than productivity growth. This is a well-known prediction from incentive-based theories of employment and has received empirical support (Lazear, 1979; Medoff and Abraham, 1981; Lazear and Moore, 1984). Existing theories have focused on wage gains within a particular job. In contrast, in our model, wage gains occur across jobs—older workers tend to be assigned to better-paying jobs that require greater incentive rents.

We conclude this subsection by noting that the observations above are consequences of optimal dynamic incentive provision at the worker level. These features of internal labor markets serve as the basis for our discussion below but do not reflect how characteristics at the firm level shape and constrain the careers of individual workers. We now turn to those issues.

6.2 Organizational Demography

In this subsection, we explore how labor-market conditions primarily affecting one level of the organization impact workers' entire career paths through endogenous organizational responses. This issue has also been the subject of study of organizational demography, which has examined, in particular, how workers' career advancement is shaped by a firm's personnel policy and organizational structure (see Stewman (1988) for a survey of the field). These studies take a firm's personnel policies and organizational structure as given. In our model, firms actively manage both their personnel policies and their organizational structure in order to maintain the motivation of their employees.

OBSERVATION 6. *Firms may adopt forced-turnover policies to create promotion opportunities.*

In the United States, prior to 1986 when this practice was outlawed, many firms put in place mandatory-retirement policies, often with the stated objective of creating promotion opportunities for the young. For example, Cappelli (2008) reports that executives at Sears put in place mandatory retirement policies "entirely to keep the lines of advancement open." The U.S. Department of Labor (1981) surveyed employers regarding this practice and summarized their results as follows: "When firms were asked for reasons for using mandatory retirement, all firms, but particularly large firms, put greatest emphasis on assuring promotional opportunities for younger workers."[2] The U.K. outlawed mandatory-retirement policies in their Employment Equality (Repeal of Retirement Age) Regulations of 2011 with the stated purpose of reducing age discrimination. In 2012, however, the Supreme Court of the U.K. granted exceptions for mandatory-retirement policies aimed at creating opportunities for younger workers.

Forced-turnover policies are not limited to mandatory-retirement programs. Many firms, including GE, Motorola, Dow Chemical, IBM, and, in the past, Microsoft, put in place "stack ranking" or "vitality curve" policies in which a fraction of workers at each level of the hierarchy is regularly dismissed. Descriptions of these policies often emphasize both the motivational effects of dismissing poor performers and that dismissing workers in higher positions creates opportunities throughout the firm.

Lazear (1979) provides a justification for mandatory-retirement policies as being part of an optimal long-term employment contract in which wage payments are backloaded in order to motivate workers, and the value of a workers' entire wage stream equals the entire stream of his contribution to profits. At termination, a worker's spot wages optimally exceed his marginal product, and therefore retirement would not be ex-post voluntary and so has to be mandated.

In our model, as in Lazear's, the NPV of workers' wage profiles determines worker participation, and the wage profile's steepness determines motivation. Because of contractual frictions, mandatory retirement allows the firm to implement, at lowest cost, steeper wage profiles. Under the optimal personnel policies, backloaded compensation implies that workers are paid less than their marginal product when they are young and more than their marginal product when they are older. However, older workers are not fired *because* their wages exceed their marginal products—their replacements, old or young, will also have to be paid wages exceeding their marginal products. Rather, older workers may be fired precisely to increase the vertical flow of workers through the organization. This result holds *even though* young workers know they may be forced out after being promoted.

As we will see in the next section, frictions in wage payments are key to our results on forced-turnover policies. In firms that make extensive use of pay-for-performance bonuses, forced-turnover policies are not necessary. Our model therefore suggests that forced-turnover policies are more likely to be prevalent in organizations in which pay is less flexible.

Even though forced-turnover policies may be optimal for some firms, the desirability of pushing workers out in order to create opportunities for others does not hold at the level of the entire

[2] We thank Michael Waldman for this quote.

economy. Yet this motive has been cited extensively as a justification for increasing the generosity of government retirement programs. For example, in the UK, the Job Release Scheme "was introduced in 1977 and was described as 'a measure which allows older workers to retire early in order to release jobs for the registered unemployed.'" (Banks, Blundell, Bozio, and Emmerson, 2010, p. 7). Changes in the skill mix notwithstanding, the argument has some intuitive appeal. After all, in the steady state, hiring at the bottom of the organization is carried out exactly to offset departures from the firm. That is, $H_1^* = d_1 N_1^* + D_2^* N_2^*$, where D_2^* represents the sum of the voluntary and involuntary departure rates at the top. All else equal, an increase in the rate of voluntary departures at the top increases hiring at the bottom, since $\partial H_1^* / \partial d_2 = N_2^* > 0$.

In our model, however, in response to an increase in the voluntary departure rate, firms optimally adjust their personnel policies and the number of workers they employ. That is,

$$\frac{dH_1^*}{dd_2} = d_1 \frac{\partial N_1^*}{\partial d_2} + D_2^* \frac{\partial N_2^*}{\partial d_2} + \frac{\partial D_2^*}{\partial d_2} N_2^*.$$

As a result, the impact of turnover should be evaluated by taking into account these changes.

OBSERVATION 7. *Increased turnover at the top may lead to less employment at the bottom and less hiring at the bottom.*

We show in Corollary 4 that an increase in the voluntary departure rate at the top of the firm can lead to a decrease in the steady-state employment level at the bottom of the firm. The reasons for this are two-fold. First, when there are insufficient separation rents, firms adopt forced-turnover policies. An increase in the rate of voluntary departures causes the firm to scale back on these forced-turnover policies (i.e., $\partial D_2^* / \partial d_2 = 0$), but firms otherwise make no other changes. In other words, firms are already able to expand opportunities for entry-level workers by increasing turnover at the top and will do so themselves when they find it profitable. Second, when there are sufficient separation rents, an increase in the voluntary departure rate causes the firm to reduce the number of positions at the top and at the bottom (i.e., $\partial N_i^* / \partial d_2 < 0$).

Empirically, the effects of increased retirement rates on the employment of younger workers is exactly the opposite of the objectives stated for the Job Release Scheme. Examining changes in the generosity of government retirement programs in twelve countries in the 20th century, several authors have shown that when government retirement programs became more (less) generous, older workers retired earlier (later), and youth and prime-age unemployment went up (down) (Gruber and Wise, 2010). Our model's flow constraint captures the intuition behind proposals like the Job Release Scheme, but it also highlights organizational reasons for why its intended outcomes might fail to materialize.

6.3 Firm Size, Span, and Workers' Careers

In this section, we contrast workers' careers in large firms to those in small firms. We group our model's predictions into static, cross-sectional observations and dynamic predictions relating to a worker's entire career. The first set of predictions speaks to the firm size–wage effect that has been widely documented in labor economics (see Oi and Idson (1999) for a survey). The second set of predictions is consistent with the findings of many disparate single-firm studies.

To think about these issues in the context of our model, we make use of the empirical pattern that larger firms tend to have larger spans (see Rushing (1966), Blau and Schoenherr (1971), Kasarda (1974), and Colombo and Delmastro (1999) for cross-sectional evidence of this pattern; see Caliendo, Monte, and Rossi-Hansberg (Forthcoming) for recent evidence related to within-firm growth). That is, we make the assumption that expansion favors activity 1, and we explore the implications of this assumption for the effects of firm size on workers' careers. As in Corollaries 5 and 6, we will consider a small firm to be one that operates in the sufficient-separation-rents region, and we consider a large firm to be one that operates in the insufficient-separation-rents region.

OBSERVATION 8. *Larger firms pay higher wages for all workers.*

A positive relationship between firm size and wages has been documented in many studies going back to at least Moore (1911) (Brown and Medoff (1989); see Oi and Idson (1999) for a review of the literature). Our explanation is closest to the efficiency-wage story that was originally posited as an explanation for the size–wage effect. Under the efficiency-wage explanation, managers have more difficulty monitoring workers in larger establishments (corresponding to a lower q_i) and therefore have to offer higher wages for all workers. But this size-monitoring explanation operates at the establishment-level, rather than the firm-level, and the evidence suggests that there is a positive relationship between firm size and wages, even controlling for establishment size.

In contrast to the efficiency-wage explanation, in our model, firms need not possess different technologies in order for a firm size–wage effect to exist. In our model, a small firm has a small span, which means that bottom workers' promotion prospects are relatively strong. As a result, the firm can offer bottom workers a lower wage, while still keeping these workers motivated. In such firms, top workers have full job security and therefore can be motivated with a relatively low wage.

Our model generates a firm size–wage effect, but it operates through a firm span–wage effect. Guadalupe and Wulf (2010) provide evidence for a firm span–wage effect, showing that an exogenous shock leads firms to increase their span and increase managers' total compensation, even after controlling for firm size. We would also predict that controlling for firm span in a regression of wages on firm size should reduce the magnitude of the size–wage effect. We are unaware of studies directly showing a size–wage effect controlling for firm span or promotion prospects.

OBSERVATION 9. *Larger firms have more of an insider bias in hiring at the top.*

The model's stark result that hiring only occurs at the bottom is due in part to worker homogeneity. The result is more continuous, however. Suppose the firm has a one-time opportunity to hire into the top job an external candidate whose incremental productivity over existing workers is Δ (i.e., the incremental increase in the NPV of future profits from hiring this worker is Δ). Since workers in the top job receive rents of v_2^*, if the firm hires this external worker, the firm has to offer him total rents of v_2^*. In contrast, if the firm instead hires into the bottom job and promotes a bottom worker, the firm will pay total rents of 0, since the expected future rents from eventually being promoted are extracted from the bottom worker. As a result, the firm will hire the external candidate into the top job only if $\Delta > v_2^*$.

The firm therefore exhibits an insider bias in hiring into higher-level positions, for which there is empirical support. For example, Huson, Malatesta, and Parrino (2004) find that outside CEOs bring about better firm performance; Agrawal, Knoeber, and Tsoulouhas (2006) find that external candidates are superior to internal candidates in observable qualities; and Oyer (2007) finds that there is an insider advantage for tenure decisions for academic economists. The mechanism we highlight here is similar to that of Chan (1996) and Waldman (2003). Our emphasis, however, is on how the degree of insider bias varies with firm size.

Since larger firms offer greater rents to top workers than smaller firms do, the insider bias in hiring at the top is greater in larger firms. There is extensive support for the idea that larger firms have more of an insider bias for hiring CEOs (Dalton and Kesner, 1983; Lauterbach and Weisberg, 1994; Parrino, 1997; Lauterbach, Vu, and Weisberg, 1999; Agrawal, Knoeber, and Tsoulouhas, 2006). More broadly, in a nationally representative sample of UK firms, recent papers find support for a positive firm size–insider bias relationship (DeVaro and Morita, 2013; Bond, 2014).

DeVaro and Morita (2013) explain this positive relationship between firm size and insider bias in hiring at the top by arguing that firms differ in the "returns to managerial capability." Firms with greater returns to managerial capability will hire more workers at the bottom and therefore operate at a larger scale. Additionally, the returns to training subordinates to become managers and the returns to promoting from within are higher in such firms. In equilibrium, there is therefore a positive correlation between firm size and insider bias in hiring at the top, driven by unobserved returns to managerial capability. In contrast, our model suggests that this positive relationship is likely to hold even among firms within narrowly defined industries, for which one might expect the returns to managerial capability to be similar. And empirically, this relationship does hold when making within-industry comparisons.

OBSERVATION 10. *Larger firms have higher starting wages and higher wages for the promoted workers, but the promotion prospects for bottom workers are worse.*

We conclude this section with a discussion of how firm characteristics affect workers' wage and career dynamics. Workers in larger firms begin their employment with higher wages but worse promotion prospects than workers in smaller firms. Their promotion prospects are worse, because larger firms optimally choose to have larger spans, which in essence creates more competition among bottom workers for promotions. To maintain incentives through promotions, larger firms also choose higher wages at the top. Nevertheless, expected future rents for bottom workers in larger firms are smaller than they are for bottom workers in smaller firms, and therefore, to keep bottom workers motivated, the firm offers higher wages at the bottom.

While we are unaware of any studies that directly examine how firm size is related to career dynamics, a number of papers have shown relationships between firm size and various aspects of wage and promotion dynamics. Taken together, their results are consistent with our predictions. In terms of starting wages, Barron, Black, and Loewenstein (1987) and Brown and Medoff (1989) find that larger firms offer higher starting wages. In terms of wages for promoted workers, there is certainly some evidence that wages at the very top are higher for larger firms—Gabaix and Landier (2008) and Tervio (2008) show that CEO compensation increases with firm size, and Garicano and Hubbard (2009) show that the pay that top lawyers receive is increasing in the size of their firms. Finally, in terms of promotion prospects, Belzil and Bognanno (2008) study the careers of over 30,000 American executives across many firms and find that the rate at which they are promoted to higher positions is negatively related to the size of their employers.

Rebitzer and Taylor (1995) study the labor market for lawyers and, in particular, focus on the relationship between firm size and various aspects of career dynamics. They find that larger law firms offer higher wages to both their associates and partners, and they interpret these findings as evidence against an efficiency-wage model—higher pay for partners in larger firms should be viewed as backloaded pay for associates, implying that associates should have lower wages in larger firms. In our model, however, bigger wage increases upon promotion do not imply lower wages at the bottom, because bottom workers' promotion prospects are lower. Indeed, this is consistent with Galanter and Palay's (1991) broad study of law firms in which they find that, "the chances of promotion to partner are accordingly lower in big firms than small firms."

To conclude this section, we note that differences in career dynamics between small firms and large firms are, of course, likely to result from many other factors, notably from differences in the quality of labor. Moreover, there are aspects of wage dynamics that cannot be explained by our model. For example, Barron, Black, and Loewenstein's (1987) findings that larger firms offer higher within-job wage growth cannot be explained by our model, because we assume that wages are constant within a job. Similarly, Baker, Gibbs, and Holmstrom's (1994b) finding that within-worker wage growth is autocorrelated is more naturally explained by models that focus on worker heterogeneity (Gibbons and Waldman, 1999a). In contrast, our model focuses on how firm-level heterogeneity, rather than worker heterogeneity, affects the career dynamics of workers. That is, differences in promotion opportunities, driven by differences in production technologies and product demand, can shed light on how workers' career patterns differ across firms.

7 Pay-for-Performance Contracts

In our main model, we made the assumption that wages are tied to activities, and we focused on steady-state analysis. These assumptions ruled out both performance pay and other more flexible arrangements such as seniority-based raises and promotions. We now expand the firm's contracting possibilities by allowing firms to use history-contingent pay-for-performance contracts and assignment policies. In order to maintain the idea that contracting is imperfect, we assume that workers are subject to a limited-liability constraint. We solve for optimal personnel policies, allowing pay and activity assignment to be history-contingent, and we show that they share most of the features of optimal personnel policies in the main model. Additionally, our results in this section shed light on the conditions under which forced-turnover policies and performance pay are part of optimal personnel policies.

Specifically, we assume that the firm pays a minimum wage $\underline{w} \geq 0$ at the beginning of each period and a performance-contingent bonus $b_t \geq 0$ at the end of each period. As in the main model, we assume that any worker who is caught shirking is terminated with probability 1, and he also receives no bonus. Indeed, this punishment gives the agent the lowest possible payoff if he shirks. If a worker is not caught shirking, the firm pays a bonus $b_t \geq 0$, which can depend on the worker's entire past employment history within the firm. For a worker in his t-th period in the firm, his employment history can be described as $h^t = (h_1, \ldots, h_t)$, where $h_s \in \{1, 2\}$, $s = 1, 2, \ldots, t$ denotes the activity to which the worker was assigned in period s. In other words, a worker's bonus in period t is determined by the function $b\left(h^t\right)$. We also allow the firm's assignment policy to depend on h^t. Denote by $p_i\left(h^t\right)$, $i = 1, 2$ the probability that a worker will be assigned to activity i in period $t+1$, given history h^t. The complementary probability $1 - p_1(h^t) - p_2(h^t)$ is then the associated forced-turnover probability. As in the main model, the firm offers the same contract to all workers, so the firm's optimal personnel policy can be described by $\{b(h^t), p_1(h^t), p_2(h^t)\}_{t=1}^{\infty}$.

Since bonuses and activity assignments can depend on the worker's entire employment history, this extension allows for a variety of personnel policies. For example, the firm can adopt seniority-based promotion policies in which each worker performs, say, activity 1 for a number of periods before being promoted to activity 2. The firm can also rotate workers among jobs. In terms of the bonus policy, the firm is not restricted to setting the same bonus level for all workers performing the same activity. A worker's bonus can depend both on the time he has spent on the activity and on the time he has been in the firm.

To describe the optimal contract, define the incentive rents for activity i as $r_i = (1-q_i)\,c_i/q_i$. As in the main model, we assume that $r_2 > r_1$, so activity 2 requires greater incentive rents, either because its associated effort costs are higher or because performance is more difficult to monitor. To simplify our discussion, we assume that the minimum wage alone is not high enough to motivate workers to exert effort in either activity.

ASSUMPTION 3. For $i = 1, 2$, $\underline{w} + r_i > (\underline{w} - c_i)/(1 - \delta(1-d_i))$.

The set of feasible bonus and assignment policies is large, so solving for the optimal contract requires a different set of techniques than the variational arguments we used in the main model. In particular, we find the optimal contract by first establishing a lower bound for labor costs required to sustain effort. We then construct a particular contract that attains this lower bound, so this contract is optimal. Proposition 2 describes the optimal contract.

PROPOSITION 2. *There is an optimal personnel policy with the following features:* (*i*) *Hiring occurs only in the bottom job, where workers perform Activity 1.* (*ii*) *There is a well-defined career path: bottom workers stay in the bottom job or are promoted. The promotion rate is constant and given by* $d_2N_2/((1-d_1)N_1)$. *Top workers perform Activity 2 and are never demoted. Workers are not fired unless they are caught shirking.* (*iii*) *The performance bonus in the top job is constant and independent of the firm's span. The performance bonus in the bottom job is also constant, and it is equal to zero if the span* N_1/N_2 *is below a threshold and is otherwise positive and increasing in the span.*

Proposition 2 shows that performance bonuses within each job are stationary—under the optimal personnel policy, pay is backloaded across jobs rather than within a job. As in the main model, an internal labor market emerges. New hires enter through a port of entry in which they perform the low-rent activity. Incumbent workers climb a job ladder, and there are no demotions. By assigning workers to the low-rent activity before promoting them to the high-rent activity, the firm uses the incentive rents for the top job to motivate both top and bottom workers. The idea of reusing rents at the top is also present in the main model and reflects optimal rent extraction at the worker level.

Firm-level characteristics, such as its span, determine optimal personnel policies. As in the main model, if the firm's span is below a threshold, separation rents created from voluntary turnover at the top, along with minimum-wage payments, are enough to motivate bottom workers. In this case, workers in the top job receive the minimal incentive bonus necessary to motivate them. Workers in the bottom job are not given performance bonuses, and their pay in each period is equal to $\underline{w}$. When there are sufficient separation rents, promotion prospects alone are strong enough to motivate the bottom workers, so no performance pay is necessary in the bottom job.

If the firm's span exceeds the threshold, we say that there are insufficient separation rents. In this case, top workers again receive the minimal incentive bonus necessary to motivate effort. Bottom workers, however, receive positive performance bonuses, and the bonus amount increases with the firm's span. When there are insufficient separation rents, incentives provided by promotions are not enough to motivate bottom workers. Therefore, the firm must adjust its personnel policy to provide additional incentives for bottom workers. Unlike the main model, however, the firm does not do so by increasing rents and putting in place forced-turnover policies at the top. Instead, it increases performance bonuses at the bottom.

The difference arises because when bonuses are feasible, they serve as a more effective way to provide additional incentives. Using bonuses involves only a monetary transfer from the firm to the workers and therefore does not affect the surplus of the relationship. In contrast, expanding promotion opportunities requires that the firm adopt forced-turnover policies, which reduce top workers' effective discount factors, limiting the firm's ability to extract rents from them.

Below, we discuss two additional observations that arise when firms are able to make use of performance pay. The first observation reinforces the theme that there are complementarities among personnel policies. In particular, it complements Observations 6 and 10 by providing conditions under which forced-turnover policies are unlikely to arise.

OBSERVATION 11. *Mandatory-retirement policies are more prevalent in settings in which pay-for-performance contracts are limited.*

We are unaware of any systematic evidence on how the prevalence of pay-for-performance relates to the adoption of mandatory-retirement policies. Casual empiricism suggests that occupations with mandatory-retirement policies are often those in which pay-for-performance contracts are rare—judges, police and military officers, government officials, and clerks. To the extent that effective pay-for-performance contracts are differentially limited by firm-level heterogeneity in monitoring technology, our model suggests that mandatory-retirement policies are more likely to be adopted in firms in which worker performance is more difficult to measure. Difficulties in implementing pay-for-performance contracts may make mandatory-retirement policies desirable.

Finally, we note in Observation 12 that even when pay-for-performance contracts are feasible, they are not necessarily used to motivate bottom workers.

OBSERVATION 12. *Bonuses are used to motivate top workers. Bottom workers are motivated by promotions, but when promotion prospects are limited, they may also be motivated by bonuses.*

Observation 12 describes how the use of bonuses varies across hierarchical levels within a firm. The firm must use bonuses to motivate top workers, since there are no further promotion prospects for them. In contrast, bottom workers are motivated primarily through promotion prospects. They receive bonuses only when promotion prospects are limited. Increased prevalence of bonus schemes higher up in firms is consistent with the evidence. For example, Gibbs (1995) examines the personnel records of a large U.S. firm and finds that the fraction of workers receiving bonuses increases monotonically with the workers' level in the firm's hierarchy.

Finally, even if pay-for-performance contracts are available, bottom workers may be motivated exclusively by promotion prospects. Our model sheds light on Baker, Jensen, and Murphy's (1988) puzzle that promotions are used as the primary incentive device in most organizations. They note that "*[t]he empirical importance of promotion-based incentives, combined with the virtual absence of pay-for-performance compensation policies, suggests that providing incentives through promotion opportunities must be less costly or more effective than providing incentives through transitory financial bonuses. This prediction is puzzling to us because promotion-based incentive schemes appear to have many disadvantages and few advantages relative to bonus-based incentive schemes.*"

If contracting were perfect in our model, promotions would provide no additional benefits over the use of bonus payments. When workers are subject to a limited-liability constraint, however, using bonus payments requires providing workers with limited-liability rents. By promising to promote bottom workers to offset voluntary turnover at the top, the firm can reduce the limited-liability rents it provides to bottom workers. As a result, the firm will fill all voluntary departures at the top with bottom workers before resorting to pay-for-performance for bottom workers.

8 Conclusion and Discussion

This paper shows that career ladders arise naturally within organizations in response to contractual imperfections. Jobs requiring lower levels of incentive rents serve as ports of entry, and workers are motivated in part by the opportunity to advance to jobs requiring, and therefore delivering, higher levels of incentive rents. When promotion opportunities are limited by the firm's organizational structure and the voluntary departure rate of its employees, firms optimally push out higher-level employees in order to keep the lines of advancement open and increase the wage growth upon promotion. Firms may also optimally alter their organizational structures, becoming more top heavy, in order to expand promotion opportunities.

The model is tractable, and we have explored many extensions. First, we can embed the model in a market setting, allowing us to study the effects of labor-market policies on the careers of workers. Preliminary analysis shows that progressive taxation, which disproportionately affects top workers, has indirect effects on bottom workers—fewer workers are hired at the bottom, but the workers who are hired have greater promotion opportunities. If firms are subject to employment protection legislation that introduces costs to adopting forced-turnover policies, optimal personnel policies involve lower wages at the top and fewer positions at the top, which in turn reduces bottom workers' promotion prospects. Finally, we demonstrate that minimum-wage policies can either increase or decrease employment in the firm. In particular, employment at the bottom of the firm can increase, since limited-liability rents can serve as a substitute for career-based incentives for bottom workers—minimum wages may lead to the proliferation of "dead-end" jobs.

Second, we can extend our model to study job-title creation in organizations. Baron and Bielby (1986)[3] show that job titles often do not correspond to different uses of the underlying production technology and are sometimes created as a way of sidestepping formal and rigid wage schedules. In our model, the firm can better extract rents from workers by creating different job titles for the same activity. Our model therefore suggests that the incentives for job-title proliferation are stronger when there are insufficient separation rents and when pay-for-performance contracts are difficult to implement. Our framework is therefore consistent with Baron and Bielby's findings that job-title proliferation is more likely to occur in large, bureaucratic organizations.

Third, we have deliberately abstracted from many of the conventional forces that have been identified in the literature, including employer learning, human-capital acquisition, and signaling, in order to emphasize the richness of empirically relevant patterns that are generated by this single force. These forces can be incorporated into the model, however, and the resulting interactions

[3]We thank Hideshi Itoh for suggesting this reference.

can generate relevant patterns. For example, by allowing worker-level heterogeneity and employer learning about worker quality, the associated optimal personnel policy is consistent with Medoff and Abraham's (1980) seniority–wage puzzle that worker tenure in a job is associated with higher wages but not higher performance. In our model, since wages in the current job and promotions serve as substitute mechanisms for motivation, a worker who has been revealed to be a particularly bad fit for promotion will have to be compensated with higher wages in the current job in order to maintain motivation. Selection would therefore account for the seniority-wage puzzle.

Fourth, the model currently considers firms with production functions requiring two activities to be performed. Allowing for more activities would generate a career ladder with more than two levels and would allow us to study how forced-turnover rates, wages, and promotion policies differ throughout the hierarchy. Our model's main mechanism suggests that wages ought to be increasing at an increasing rate as workers climb the career ladder—larger wage increases at higher levels provide stronger incentives than corresponding wage increases at lower levels (Rosen, 1986).

Moreover, by incorporating both a three-tier hierarchy and learning about worker quality into our model, we may be able to shed light on empirical findings that are especially difficult to reconcile with existing theories. For example, Baker, Gibbs, and Holmstrom (1994b) find overlap in the distribution of wages for different levels of the hierarchy—the highest-paid worker at one level receives a higher wage than the lowest-paid worker at the next level. Overlapping wage distributions could arise in our model, because the highest-paid worker at one level likely has the lowest promotion prospects. In contrast, the lowest-paid worker at the next level is likely to have been recently promoted and still has strong prospects for being promoted to the subsequent level. The prospect of further promotions implies that such a worker is willing to exert effort even if his current wage is below the wage of the highest-paid worker at the level below.

Fifth, we have assumed that workers' outside options are exogenous. In human-capital-intensive industries, many firms adopt practices that intentionally or unintentionally increase workers' outside options. For example, firms often provide training that increases a worker's general human capital (Becker, 1975), and many firms offer outplacement services for workers whose jobs are eliminated. Some firms, especially in the management-consulting industry, actively invest in placing workers who are forced to leave because of up-or-out policies. A McKinsey insider commented that, "if international companies stopped recruiting former McKinsey staff, it could clog the 'up or out' refining process." In the context of our model, if training increases the outside option of top workers by more than it increases the outside option for bottom workers, it increases the value that workers place on being promoted and can therefore help reduce distortions in the firm's organizational structure and wages. The model therefore suggests important interactions between firms' training and outplacement policies and the rest of their personnel policies.

Finally, we have focused on a steady-state analysis. The firm's size and organizational structure therefore do not change over time. Allowing for a nonstationary environment would allow us to examine how firm growth interacts with a firm's optimal personnel policies. It seems natural to think that a firm experiencing a higher growth rate can better rely on promotion incentives to motivate their workers. At some point, however, high-growth firms mature, and their growth slows. Jensen (1986a,b) argues that in slowly growing firms, managers may spend resources to pursue unprofitable growth. Understanding why firms pursue seemingly unprofitable growth and understanding how firms change their personnel policies in response to a slowdown in growth are intriguing theoretical questions with important practical implications.

9 Appendix

LEMMA 0. *A firm maximizing its profits separately over the two activities chooses wages* $w_i^{pc} = c_i + (1-(1-d_i)\,\delta)\,R_i$ *to provide rents* $v_i^{pc} = R_i$ *to each worker performing activity* $i = 1, 2$. *The firm hires* $H_i^{pc} = (1-d_i)\,N_i^{pc}$ *workers, where* $F_i\left(N_i^{pc}, N_j^{pc}\right) = w_i^{pc} > c_i$.

PROOF OF LEMMA 0. For activity i, the firm will choose a wage that ensures the incentive-compatibility constraint holds with equality:

$$v_i = w_i - c_i + (1-d_i)\,\delta v_i = w_i + (1-q_i)\,(1-d_i)\,\delta v_i,$$

which gives us

$$v_i^{pc} = c_i / \left(q_i\,(1-d_i)\,\delta\right),$$

which is the desired expression.■

We next establish Proposition 1', which describes the optimal personnel policy for any organizational structure (N_1, N_2). Proposition 1 in the text is the special case when $N_1 \geq N_2$. First, we establish Lemmas 1-3, which hold if $v_1^* \leq v_2^*$. In the proof of Proposition 1', we show that when $N_1 \geq N_2$, indeed $v_1^* \leq v_2^*$. We establish an analogous set of conditions for the case when $N_1 < N_2$.

LEMMA 1. *Assume* $v_1^* \leq v_2^*$. *Then* $H_2^* = 0$.

PROOF OF LEMMA 1. Denote the mass of incumbent workers in activity i by $M_i = (1-d_i)\,N_i$, and substitute $(PK-1)$ and $(PK-2)$ into the firm's wage bill:

$$\begin{aligned}
W &= w_1 N_1 + w_2 N_2 \\
&= (v_1 + c_1 - \delta\,(1-d_1)\,(p_{11}v_1 + p_{12}v_2))\,N_1 + (v_2 + c_2 - \delta\,(1-d_2)\,(p_{21}v_1 + p_{22}v_2))\,N_2 \\
&= c_1 N_1 + c_2 N_2 + v_1\,(N_1 - \delta\,((1-d_1)\,p_{11}N_1 + (1-d_2)\,p_{21}N_2)) \\
&\quad + v_2\,(N_2 - \delta\,((1-d_1)\,p_{12}N_1 + (1-d_2)\,p_{22}N_2)) \\
&= N_1 c_1 + N_2 c_2 + v_1\,((1-\delta)\,N_1 + \delta H_1) + v_2\,((1-\delta)\,N_2 + \delta H_2),
\end{aligned}$$

where the third equality holds by $(FL-1)$ and $(FL-2)$. Minimizing the wage bill, when the flow constraint has been substituted in, is therefore equivalent to minimizing

$$v_1\,((1-\delta)\,N_1 + \delta H_1) + v_2\,((1-\delta)\,N_2 + \delta H_2).$$

We now show that $H_2^* = 0$. Consider an optimal W^* with $H_2^* > 0$ and $v_2^* \geq v_1^*$. Since $M_1 + M_2 > N_2$, either $p_{12}^* < 1$ or $p_{22}^* < 1$. In the first case, consider a perturbation in which fewer workers are hired into activity 2, and instead, they are hired into activity 1, and a larger fraction of activity-1 workers are reassigned to activity 2 in the next period. That is, let $\tilde{H}_1 = H_1^* + M_1\varepsilon$, $\tilde{H}_2 = H_2^* - M_1\varepsilon$, $\tilde{p}_{11} = p_{11}^* - \varepsilon$, and $\tilde{p}_{12} = p_{12}^* + \varepsilon$. In the second case, consider a perturbation in which again, fewer workers are hired into activity 2, and instead, they are hired into activity 1, and a larger fraction of activity-2 workers are assigned to activity 2 again in the next period. That is, let $\tilde{H}_1 = H_1^* + M_2\varepsilon$, $\tilde{H}_2 = H_2^* - M_2\varepsilon$, $\tilde{p}_{21} = p_{21}^* - \varepsilon$, and $\tilde{p}_{22} = p_{22}^* + \varepsilon$. Under either of these perturbations,

$$\tilde{W}_j = W - \delta M_j \varepsilon\,(v_2^* - v_1^*) \leq W^*.$$

If $v_2^* > v_1^*$ is strict, the above inequality shows that the original personnel policy cannot be optimal. If $v_2^* = v_1^*$, the perturbation does not affect the wage bill, and indeed, the perturbation can be continued up to the point where $\tilde{H}_2 = 0$. Therefore, $H_2^* = 0$.■

LEMMA 2. *Assume* $v_1^* \leq v_2^*$ *and* $(1-d_1) N_1 R_1 \leq d_2 N_2 R_2$*. Then* $v_1^* = 0$*,* $v_2^* = R_2$*,* $p_{21}^* = 0$ *and* $p_{i1}^* + p_{i2}^* = 1$ *for* $i = 1, 2$*.*

PROOF OF LEMMA 2. We first show that $v_2^* = R_2$ and $v_1^* = 0$. By $(IC-2)$, it must be the case that $v_2 \geq R_2$. Note that $v_1 \geq 0$, by $(IR-1)$. Therefore, if $(v_1, v_2) = (0, R_2)$ is attainable, it will minimize the wage bill. Since p_{ij} does not enter the cost function directly, it suffices to show that there exists an assignment matrix $\mathbf{P}$ such that under $\mathbf{P}$, $(v_1, v_2) = (0, R_2)$ satisfies all the constraints. This is indeed the case, since we can set $p_{22}^* = 1$ and

$$p_{12}^* = (N_2 - M_2) / M_1 \leq 1,$$

so that $(IC-1)$ becomes $p_{12}^* R_2 \geq R_1$, and is therefore satisfied since $d_2 N_2 R_2 \geq (1-d_1) N_1 R_1$. Therefore, $(v_1^*, v_2^*) = (0, R_2)$ and $v_1^* < v_2^*$ is indeed the case. Clearly, $p_{22}^* = 1$ implies that there are no demotions (i.e., $p_{21}^* = 0$) and no forced turnover at the top.

Next, we show that there is no forced turnover at the bottom. Given $H_2^* = 0$ from Lemma 1 and $p_{22}^* = 1$, which we just showed, we can add up $(FL-1)$ and $(FL-2)$ to obtain

$$(p_{11}^* + p_{12}^*) M_1 + M_2 + H_1^* = N_1 + N_2,$$

which implies that

$$H_1^* \geq (1 - p_{11}^* - p_{12}^*) M_1 + N_1 - M_1.$$

Assume $p_{11}^* + p_{12}^* < 1$, and consider the following perturbation. Let $\tilde{H}_1 = H_1^* - M_1 \varepsilon$ and $\tilde{p}_{11} = p_{11}^* + \varepsilon$ for some $\varepsilon > 0$. The flow constraint is still satisfied, since

$$\tilde{p}_{11} M_1 + p_{21}^* M_2 + \tilde{H}_1 = N_1,$$

and this increase in p_{11}^* does not affect $(IC-1)$. All other constraints remain satisfied. This perturbation therefore satisfies all the constraints and weakly reduces the wage bill. We can therefore continue this perturbation until $p_{11}^* + p_{12}^*$, where $\tilde{H}_1 > 0$ is still true. Therefore, it is optimal to set $p_{11}^* + p_{12}^* = 1$, and there is no forced turnover at the bottom.■

LEMMA 3. *Assume* $v_1^* \leq v_2^*$ *and* $(1-d_1) N_1 R_1 > d_2 N_2 R_2$*. Then* $v_1^* = 0$*,* $v_2^* > R_2$*,* $p_{21}^* = 0$*,* $p_{22}^* < 1$*, and* $p_{11}^* + p_{12}^* = 1$*.*

PROOF OF LEMMA 3. Define $\Delta_i \equiv p_{i1} v_1 + p_{i2} v_i - R_i$ to be the excess rents provided to workers performing activity i. We can write

$$\begin{aligned} M_1 \Delta_1 + M_2 \Delta_2 &= (M_1 p_{11} + M_2 p_{21}) v_1 + (M_{12} p_{12} + M_2 p_{22}) v_2 - M_1 R_1 - M_2 R_2 \\ &= (N_1 - H_1) v_1 + (N_2 - H_2) v_2 - M_1 R_1 - M_2 R_2, \end{aligned}$$

where the second equality uses the flow constraints $(FL-1)$ and $(FL-2)$. This equality allows us to rewrite the objective function as

$$\begin{aligned} W &= N_1 c_1 + N_2 c_2 + ((1-\delta) N_1 + \delta H_1) v_1 + ((1-\delta) N_2 + \delta H_2) v_2 \\ &= N_1 c_1 + N_2 c_2 + (1-\delta)(M_1 R_1 + M_2 R_2) + H_1 v_1 + H_2 v_2 + (1-\delta)(M_1 \Delta_1 + M_2 \Delta_2). \end{aligned}$$

Note that this implies that

$$W \geq N_1 c_1 + N_2 c_2 + (1-\delta)\left(M_1 R_1 + M_2 R_2\right).$$

Therefore, if $v_1 = 0$ and $\Delta_i = 0$ are attainable (and we can choose $H_2 = 0$), the labor-cost function will be minimized.

When $d_2 N_2 R_2 < (1-d_1) N_1 R_1$, we can confirm that $v_1 = 0$ and $\Delta_i = 0$ are attainable. We can use the flow constraints $(FL-1)$ and $(FL-2)$ to show that $\Delta_i = 0$ implies that

$$v_2^* = \frac{R_1 M_1 + R_2 M_2}{N_2} > R_2 > 0.$$

The corresponding assignment probabilities are feasible, since

$$\begin{aligned} p_{12}^* &= \frac{R_1 N_2}{R_1 M_1 + R_2 M_2} \in (0,1), \\ p_{22}^* &= \frac{R_2 N_2}{R_1 M_1 + R_2 M_2} \in (0,1). \end{aligned}$$

Therefore, the optimal solution is

$$v_1^* = 0,\ v_2^* = \frac{R_1 M_1 + R_2 M_2}{N_2}.$$

We next show that no demotions (i.e., $p_{21}^* = 0$) can be part of the optimum. Since $v_1^* = 0$, if $p_{21}^* > 0$, we can decrease p_{21}^* by ε. Define $\tilde{p}_{21} = p_{21}^* - \varepsilon$ and $\tilde{H}_1 = H_1^* + M_2 \varepsilon$ for some $\varepsilon > 0$. The flow constraint is still satisfied, since

$$p_{11}^* M_1 + \tilde{p}_{21} M_2 + \tilde{H}_1 = N_1.$$

Further, decreasing p_{21}^* does not affect $(IC-1)$ given that $v_1^* = 0$. We can continue this perturbation until $\tilde{p}_{21} = 0$. When $p_{21}^* = 0$, we have that $p_{21}^* + p_{22}^* = p_{22}^* < 1$, which implies that there is forced turnover at the top.

Finally, we can use the same logic as in the proof of Lemma 2 to show that $p_{11}^* + p_{12}^* = 1$ can be part of an optimum, in which case there is no voluntary turnover at the bottom.■

PROPOSITION 1'. *Define* $MR = M_1 R_1 + M_2 R_2$. *The following is true:*

(i) If $N_1 R_2 \leq MR$ *and* $N_2 R_2 \leq MR$*, then* $v_1^* = 0, v_2^* = MR/N_2, p_{11}^* = 1 - p_{12}^*, p_{12}^* = N_2 R_1/MR, p_{21}^* = 0, p_{22}^* = N_2 R_2/MR, H_1^* = N_1 d_1 + N_1 (1-d_1)(1-p_{12}^*)$*, and* $H_2^* = 0$.

(ii) If $N_1 R_2 > MR$ *and* $N_2 R_2 \leq MR$*, then* $v_1^* = 0, v_2^* = MR/N_2, p_{11}^* = 1 - p_{12}^*, p_{12}^* = N_2 R_1/MR, p_{21}^* = 0, p_{22}^* = N_2 R_2/MR, H_1^* = N_1 d_1 + N_1 (1-d_1)(1-p_{12}^*)$*, and* $H_2^* = 0$.

(iii) If $N_1 R_2 \leq MR$ *and* $N_2 R_2 > MR$*, then* $v_1^* = MR/N_1, v_2^* = 0, p_{11}^* = N_1 R_1/MR, p_{12}^* = 0, p_{21}^* = N_1 R_2/MR, p_{22}^* = 1 - p_{21}^*, H_1^* = 0$*, and* $H_2^* = N_2 d_2 + N_2 (1-d_2)(1-p_{21}^*)$.

$(iv-a)$ If $N_1 R_2 \geq N_2 R_2 \geq MR$*, then* $v_1^* = 0, v_2^* = R_2, p_{11}^* = 1 - p_{12}^*, p_{12}^* = d_2 N_2/(1-d_1) N_1, p_{21}^* = 0, p_{22}^* = 1, H_1^* = d_1 N_1 + d_2 N_2, H_2^* = 0$.

$(iv-b)$ If $N_2 R_2 \geq N_1 R_2 \geq MR$*, then* $v_1^* = R_2, v_2^* = 0, p_{11}^* = (N_1 - M_2)/M_1, p_{12}^* = 0, p_{21}^* = 1, p_{22}^* = 0, H_1^* = 0, H_2^* = N_2$.

PROOF OF PROPOSITION 1': Lemmas 1-3 establish the optimal personnel policy when $v_1^* \leq v_2^*$.

Now, we consider the case in which $v_1^* > v_2^*$ and describe the conditions under which this is true. To do so, it is useful to establish two lower bounds for the wage bill. First, recall from the proof of Lemma 3 that one lower bound for the wage bill (that holds whether $v_1^* \leq v_2^*$ or $v_1^* > v_2^*$) is:

$$W \geq N_1 c_1 + N_2 c_2 + (1-\delta)(M_1 R_1 + M_2 R_2).$$

To establish the second lower bound, notice that if $v_1^* > v_2^*$, $(IC-1)$ and $(IC-2)$ together imply that $v_1^* \geq R_2$. As a result, we have

$$\begin{aligned} W &= N_1 c_1 + N_2 c_2 + v_1((1-\delta)N_1 + \delta H_1) + v_2((1-\delta)N_2 + \delta H_2) \\ &\geq N_1 c_1 + N_2 c_2 + (1-\delta)N_1 R_2, \end{aligned}$$

where the expression for the wage bill (that holds whether $v_1^* \leq v_2^*$ or $v_1^* > v_2^*$) is obtained in Lemma 1. This is our second lower bound.

It follows that when $v_1^* > v_2^*$,

$$W \geq N_1 c_1 + N_2 c_2 + (1-\delta)\max\{M_1 R_1 + M_2 R_2, N_1 R_2\}$$

Next, we establish a personnel policy with a wage bill that reaches the lower bound. Therefore, if $v_1^* > v_2^*$, this personnel policy is optimal.

First consider the case when $N_1 R_2 \leq M_1 R_1 + M_2 R_2$. In this case, let $v_1^* = (M_1 R_1 + M_2 R_2)/N_1$ and $v_2^* = 0$. Let $p_{11}^* = N_1 R_1/(M_1 R_1 + M_2 R_2)$, $p_{12}^* = 0$, $p_{21}^* = N_1 R_2/(M_1 R_1 + M_2 R_2)$, and $p_{22}^* = 1 - p_{21}^*$. Notice that these probabilities are less than 1 because $M_1 R_1 + M_2 R_2 \leq N_1 R_2$. In addition, $(IC-1)$, $(IC-2)$, $(IR-1)$, and $(IR-2)$ are satisfied given the set of choices. Now let $H_1^* = 0$ and $H_2^* = N_2 d_2 + N_2(1-d_2)p_{21}^*$. The flow constraints are satisfied. Finally, the wage bill associated with this policy is given by $N_1 c_1 + N_2 c_2 + (1-\delta)(M_1 R_1 + M_2 R_2)$, which reaches the lower bound.

Next, consider the case with $N_1 R_2 > M_1 R_1 + M_2 R_2$. In this case, let $v_1^* = R_2$ and $v_2^* = 0$. Let $p_{11}^* = (N_1 - M_2)/M_1$, $p_{12}^* = 1 - p_{11}^*$, $p_{21}^* = 1$, and $p_{22}^* = 0$. Note that because $N_1 R_2 > M_1 R_1 + M_2 R_2$, we have $N_1 R_2 > M_2 R_2$, and therefore $p_{11}^* = (N_1 - M_2)/M_1 > 0$. Similarly, $d_1 + d_2 < 1$ guarantees that $p_{11}^* < 1$. It follows that these probabilities are between 0 and 1. In addition, $(IC-1)$, $(IC-2)$, $(IR-1)$, and $(IR-2)$ are satisfied given the set of choices. Now let $H_1^* = 0$ and $H_2^* = N_2$. The flow constraints are satisfied. Finally, the wage bill associated with this policy is given by $N_1 c_1 + N_2 c_2 + (1-\delta)N_1 R_2$, which reaches the lower bound.

Combining these two cases, we obtain the result that when $v_1^* > v_2^*$, the total wage bill is given by

$$W|_{v_1^* > v_2^*} = \begin{cases} N_1 c_1 + N_2 c_2 + (1-\delta)(M_1 R_1 + M_2 R_2) & if \quad N_1 R_2 \leq M_1 R_1 + M_2 R_2 \\ N_1 c_1 + N_2 c_2 + (1-\delta) N_1 R_2 & if \quad N_1 R_2 > M_1 R_1 + M_2 R_2, \end{cases}$$

and this is generated by the personnel policy described in the proposition.

Next, to characterize the optimal personnel policy, recall from Lemma 1 to 3 that when $v_1^* \leq v_2^*$, the total wage bill is given by

$$W|_{v_1^* \leq v_2^*} = \begin{cases} N_1 c_1 + N_2 c_2 + (1-\delta)(M_1 R_1 + M_2 R_2) & if \quad N_2 R_2 \leq M_1 R_1 + M_2 R_2 \\ N_1 c_1 + N_2 c_2 + (1-\delta) N_2 R_2 & if \quad N_2 R_2 > M_1 R_1 + M_2 R_2. \end{cases}$$

Now combining $W|_{v_1^*>v_2^*}$ and $W|_{v_1^*\leq v_2^*}$, we see that if either N_2R_2 or N_1R_2 is smaller than $M_1R_1 + M_2R_2$, the total wage bill is minimized at $W^* = N_1c_1 + N_2c_2 + (1-\delta)(M_1R_1 + M_2R_2)$. When both N_2R_2 and N_1R_2 are greater than $M_1R_1 + M_2R_2$, the minimized wage bill is given by $W^* = N_1c_1 + N_2c_2 + (1-\delta)\min\{N_1, N_2\} R_2$, and the optimal personnel policy places the activity with fewer positions on top. This describes the minimized wage bill for any organizational structure (N_1, N_2).

Finally, note that when $N_1 \geq N_2$, we always have $N_1R_2 \geq N_2R_2$. As a result, it is always optimal to put activity 2 on top. This establishes Proposition 1 in the main text.■

COROLLARY 1'. *Define $MR = M_1R_1 + M_2R_2$. The following is true:*

(i) If $N_1R_2 \leq MR$ and $N_2R_2 \leq MR$, the labor-cost function is

$$W(N_1, N_2) = N_1c_1 + N_2c_2 + (1-\delta)(M_1R_1 + M_2R_2)$$

(ii) If $N_1R_2 > MR$ and $N_2R_2 \leq MR$, the labor-cost function is

$$W(N_1, N_2) = N_1c_1 + N_2c_2 + (1-\delta)(M_1R_1 + M_2R_2)$$

(iii) If $N_1R_2 \leq MR$ and $N_2R_2 > MR$, the labor-cost function is

$$W(N_1, N_2) = N_1c_1 + N_2c_2 + (1-\delta)(M_1R_1 + M_2R_2)$$

$(iv-a)$ If $N_1R_2 \geq N_2R_2 \geq MR$, the labor-cost function is

$$W(N_1, N_2) = N_1c_1 + N_2c_2 + (1-\delta)N_2R_2$$

$(iv-b)$ If $N_2R_2 \geq N_1R_2 \geq MR$, the labor-cost function is

$$W(N_1, N_2) = N_1c_1 + N_2c_2 + (1-\delta)N_1R_2$$

PROOF OF COROLLARY 1'. Follows immediately from the proof of Proposition 1'.■

COROLLARY 2. *Given a production level y, the optimal organizational structure satisfies $N_1^*(y) > N_2^*(y)$, and the following conditions hold.*

(i) When $N_1^*(y) < \kappa N_2^*(y)$, the marginal rate of technical substitution is given by

$$MRTS(N_1^*, N_2^*) = \frac{\delta - \delta d_2}{1 - \delta d_2}\frac{c_1}{c_2}.$$

(ii) When $N_1^*(y) = \kappa N_2^*(y)$, the marginal rate of technical substitution satisfies

$$MRTS(N_1^*, N_2^*) \in \left[\frac{\delta - \delta d_2}{1 - \delta d_2}\frac{c_1}{c_2}, \frac{c_1}{c_2}\right].$$

(iii) When $N_1^*(y) > \kappa N_2^*(y)$, the marginal rate of technical substitution is given by

$$MRTS(N_1^*, N_2^*) = \frac{c_1}{c_2}.$$

PROOF OF COROLLARY 2. For the first claim that $N_1^*(y) > N_2^*(y)$, notice that if If $MRTS(N_1, N_2) > W_1^-(N_1, N_2)/W_2^-(N_1, N_2)$, where W_i^- is the left derivative of the labor-cost function, then the firm can always reduce N_2 and increase N_1, holding the production level constant, and reduce the wage bill. When $N_2 \geq N_1$, we know that $MRTS(N_1, N_2) \geq MRTS(N, N) > (c_1 + R_2)/c_2 \geq W_1^-(N_1, N_2)/W_2^-(N_1, N_2)$. As a result, it can never be optimal to have $N_2^*(y) \geq N_1^*(y)$.

For the remaining three results, note that the problem is a convex optimization program since for all y and all W, the upper contour set of production at y and the lower contour set for costs at W are convex sets. Define the function

$$\xi_W(N_1) = \{N_2 : W(N_1, N_2) = W\},$$

and define the superdifferential of ξ at W to be the set of all vectors (in (N_1, N_2) space) tangent to ξ at W:

$$\partial\xi_W = \left\{\tilde{w} = (\tilde{w}_1, \tilde{w}_2) : \tilde{w} \cdot \left(W' - W\right) \geq \xi\left(W'\right) - \xi(W) \text{ for all } W' \in \mathbb{R}_+\right\}.$$

Since the production function is twice continuously differentiable, the subdifferential of each isoquant at y is a singleton and is equal to the vector of marginal revenue products (Pf_1, Pf_2) evaluated at $(N_1(y), N_2(y))$, where $f(N_1(y), N_2(y)) = y$.

Given output y, the cost-minimization program is therefore a convex optimization program with a nondifferentiable constraint set. The associated optimality conditions ensure that for some W,

$$MRTS(N_1^*(y), N_2^*(y)) = \tilde{w}_1/\tilde{w}_2 \text{ for some } (\tilde{w}_1, \tilde{w}_2) \in \partial\xi(W).$$

In each of the three regions identified in the statement of Corollary 2, these optimality conditions correspond to the associated condition stated in the Corollary.■

COROLLARY 3. *At the optimum, $w_1^* < MRP_1^*$ and $w_2^* > MRP_2^*$.*

PROOF OF COROLLARY 3. By Corollaries 1 and 2, if at the optimum, $N_1^* > \kappa N_2^*$ (i.e., there are sufficient separation rents), we have that

$$MRP_1^* = c_1 > c_1 - c_1\kappa\frac{N_2^*}{N_1^*} = w_1^*$$

and

$$MRP_2^* = \frac{1 - \delta d_2}{1 - d_2}\frac{1}{\delta}c_2 < \frac{1}{1 - d_2}\frac{1}{\delta}c_2 = w_2^*.$$

If at the optimum, $N_1^* < \kappa N_2^*$ (i.e., there are insufficient separation rents), we have that

$$MRP_1^* = \frac{1}{\delta}c_1 > 0 = w_1^*$$

and

$$MRP_2^* = \frac{1}{\delta}c_2 < \frac{1}{\delta}c_1\frac{N_1^*}{N_2^*} + \frac{1}{\delta}c_2 = w_2^*.$$

Next, note that along the ray $N_1 = \kappa N_2$, recall from Corollary 1 that $w_1^* = 0$ and $w_2^* = (1/(1 - d_2))(c_2/\delta)$. Since $MRP_1 = Pf_1(\kappa N_2, N_2) > 0$, we therefore have $MRP_1 > 0 = w_1^*$.

Further, along the ray $N_1 = \kappa N_2$, the objective becomes

$$\max_{N_2} Pf\left(\kappa N_2, N_2\right) - w_2^* N_2,$$

so the first-order conditions can be written as $MRP_2 = w_2^* - \kappa MRP_1^* < w_2^*$.■

COROLLARY 4. *The following is true: (i) if $N_1^*/N_2^* > \kappa$, then $dN_1^*/dd_2 = dN_2^*/dd_2 = 0$; (ii) if $N_1^*/N_2^* = \kappa$, then $dN_2^*/dd_2 < 0$; and (iii) if $N_1^*/N_2^* < \kappa$, then $dN_1^*/dd_2 \leq 0$ and $dN_2^*/dd_2 < 0$.*

PROOF OF COROLLARY 4. Suppose $N_1^*/N_2^* > \kappa$. Then an increase in d_2 will exactly be offset by a corresponding decrease in the forced-turnover rate. This increase in d_2 therefore has no effect on either the production side or the total wage bill, and therefore $dN_1^*/dd_2 = dN_2^*/dd_2 = 0$.

Next, suppose $N_1^*/N_2^* = \kappa$. Then

$$\frac{d}{dd_2}\left(\frac{N_1^*}{N_2^*}\right) = \frac{d\kappa}{dd_2} = \frac{\kappa}{d_2\left(1-d_2\right)} > 0$$

and therefore

$$\frac{d}{dd_2}\left(\frac{N_1^*}{N_2^*}\right) = \kappa\left(\frac{1}{N_1^*}\frac{dN_1^*}{dd_2} - \frac{1}{N_2^*}\frac{dN_2^*}{dd_2}\right) > 0,$$

so that $\kappa dN_2^*/dd_2 < dN_1^*/dd_2$. If $dN_1^*/dd_2 < 0$, then so is dN_2^*/dd_2. We still need to show that when $dN_2^*/dd_2 > 0$, we have $dN_1^*/dd_2 < 0$. When $N_1 = \kappa N_2$, the wage bill is given by

$$W\left(\kappa N_2, N_2\right) = \frac{1}{\delta}\frac{1}{1-d_2}c_2 N_2,$$

which is increasing in d_2. Therefore, an increase in d_2 increases the marginal cost of production and therefore must result in lower total production, so if $dN_1^*/dd_2 > 0$, it must be that $dN_2^*/dd_2 < 0$. Therefore, $dN_2^*/dd_2 < 0$.

Finally, suppose $N_1^*/N_2^* < \kappa$. Then by the optimality conditions in Corollary 2,

$$\begin{aligned} Pf_1\left(N_1^*, N_2^*\right) &= c_1 \\ Pf_2\left(N_1^*, N_2^*\right) &= \frac{1-\delta d_2}{1-d_2}\frac{1}{\delta}c_2. \end{aligned}$$

Implicitly differentiating the first optimality condition, we can solve for dN_1^*/dd_2:

$$\frac{dN_1^*}{dd_2} = \frac{f_{12}}{\left(-f_{11}\right)}\frac{dN_2^*}{dd_2},$$

and implicitly differentiating the second optimality condition and solving for dN_2^*/dd_2, we get

$$\frac{dN_2^*}{dd_2} = f_{11}\frac{1}{\left(1-d_2\right)^2}\frac{1-\delta}{\delta}\frac{c_2}{P\left(f_{11}f_{22} - f_{12}^2\right)}.$$

Since f is concave, $f_{11} < 0$ and $f_{11}f_{22} - f_{12}^2 > 0$, so $dN_2^*/dd_2 < 0$. Since $f_{12} \geq 0$, $dN_1^*/dd_2 < 0$.■

COROLLARY 5. *Suppose production expansion favors activity 1. Then there exists two cutoffs, y_1 and y_2, such that the following is true: (i) if $y^* < y_1$, the firm's optimal span is $N_1^*/N_2^* < \kappa$;*

(ii) if $y^* \in [y_1, y_2]$, the firm's optimal span is $N_1^*/N_2^* = \kappa$; and (iii) if $y^* > y_2$, the firm's optimal span is $N_1^*/N_2^* > \kappa$.

PROOF OF COROLLARY 5. Denote y^* to be the optimal output level. Define the class of functions

$$\xi_W(N_1) = \{N_2 : W(N_1, N_2) = W\}.$$

From Corollary 2, optimal production always satisfies $N_1^*(y^*) > N_2^*(y^*)$. Also, optimal production will always occur at a point $\left(N_1^*(y^*), \xi_{W(y^*)}(N_1^*(y^*))\right)$ for some $W(y^*)$. Note that for each W, the function $\xi_W(N_1)$ is decreasing, piecewise-linear, and concave in N_1. Define the set

$$PEP \equiv \left\{(N_1, N_2) : N_1 = N_1^*(y^*), N_2^* = \xi_{W(y^*)}(N_1^*(y^*)) \text{ for some } y^*, W(y^*)\right\}.$$

PEP is the production-expansion path for the production function f and the labor-cost function $W(N_1, N_2)$.

Suppose $\lim_{y^* \to 0} MRTS\left(N_1^*(y^*), \xi_{W(y^*)}(N_1^*(y^*))\right) = (c_1/(c_2 + (1-\delta)R_2))$. Along the ray $(\kappa N_2, N_2)$, there exists some $\hat{N}_2$ such that for all $N_2 < \hat{N}_2$,

$$MRTS(\kappa N_2, N_2) \leq \frac{\delta - \delta d_2}{1 - \delta d_2}\frac{c_1}{c_2}.$$

Take $N_{2,1}$ to be the supremum over all such $\hat{N}_2$. It must therefore be the case that

$$MRTS(\kappa N_{2,1}, N_{2,1}) = \frac{\delta - \delta d_2}{1 - \delta d_2}\frac{c_1}{c_2}.$$

Define y_1 to be the supremum over all y such that

$$MRTS\left(\kappa N_2^*(y), \xi_{W(y)}(\kappa N_2^*(y))\right) = \frac{\delta - \delta d_2}{1 - \delta d_2}\frac{c_1}{c_2}.$$

For all $y^* \leq y_1$, optimal production satisfies

$$MRTS\left(N_1^*(y^*), \xi_{W(y^*)}(N_1^*(y^*))\right) = \frac{\delta - \delta d_2}{1 - \delta d_2}\frac{c_1}{c_2},$$

and the firm's optimal span, $N_1^*(y^*)/\xi_{W(y^*)}(N_1^*(y^*))$ is strictly increasing in y^* in this region.

Next, along the ray $(\kappa N_2, N_2)$, there exists some $\hat{N}_2$ such that for all $N_2 > \hat{N}_2$,

$$MRTS(\kappa N_2, N_2) \geq \frac{c_1}{c_2}.$$

Take $N_{2,2}$ to be the infimum over all such $\hat{N}_2$. It must therefore be the case that

$$MRTS(\kappa N_{2,2}, N_{2,2}) = \frac{c_1}{c_2}.$$

Define y_2 to be the infimum over all y such that

$$MRTS\left(\kappa N_2^*\left(y\right), \xi_{W(y)}\left(\kappa N_2^*\left(y\right)\right)\right) = \frac{c_1}{c_2}.$$

Because $c_1/c_2 > (c_1/c_2) \cdot ((\delta - \delta d_2) / (1 - \delta d_2))$, we must have that $y_2 > y_1$. Further, for all $y_1 \leq y^* \leq y_2$, optimal production satisfies

$$MRTS\left(N_1^*\left(y^*\right), \xi_{W(y^*)}\left(N_1^*\left(y\right)\right)\right) \in \left[\frac{\delta - \delta d_2}{1 - \delta d_2}\frac{c_1}{c_2}, \frac{c_1}{c_2}\right],$$

and the firm's optimal span is $N_1^*\left(y^*\right) / \xi_{W(y^*)}\left(N_1^*\left(y^*\right)\right) = \kappa$.

Finally for all $y^* > y_2$, optimal production must satisfy

$$MRTS\left(N_1^*\left(y^*\right), \xi_{W(y^*)}\left(N_1^*\left(y^*\right)\right)\right) = \frac{c_1}{c_2},$$

and the firm's optimal span, $N_1^*\left(y^*\right) / \xi_{W(y^*)}\left(N_1^*\left(y^*\right)\right)$ is strictly increasing in y^* in this region.■

COROLLARY 6. *Suppose production expansion favors activity 1. Wages are higher in both positions at large firms relative to small firms. Promotion probabilities are higher at small firms, and large firms put in place forced-turnover policies.*

PROOF OF COROLLARY 6.For small firms, $N_1^* > \kappa N_2^*$, so by Corollary 1, $w_1^{*S} = c_1 - c_1 \kappa N_2/N_1$ and $w_2^{*S} = c_2/\left(\left(1 - d_2\right)\delta\right)$. Further, the promotion rate for bottom workers is $p_{12}^{*S} = d_2 N_2/\left(\left(1 - d_1\right) N_1\right)$, and there is no forced turnover at the top, so that $p_{22}^{*S} = 1$ and $v_2^{*S} = R_2$

For large firms, $N_1^* < \kappa N_2^*$, so by Corollary 1, $w_1^{*L} = 0$ and $w_2^{*L} = \left(c_1 N_1 + c_2 N_2\right)/\left(\delta N_2\right)$. Further, by Lemma 3, we have

$$p_{12}^{*L} = \frac{R_1 N_2}{R_1 M_1 + R_2 M_2},$$

and there is forced turnover so $p_{22}^{*L} < 1$ and $v_2^{*L} > R_2$.

Putting these results together, we have $w_1^{*L} > w_1^{*S}$, $w_2^{*L} > w_2^{*S}$, $p_{12}^{*L} < p_{12}^{*S}$, $p_{22}^{*L} < p_{22}^{*S}$, and $v_2^{*L} > v_2^{*S}$.■

PROPOSITION 2. *There is an optimal personnel policy with the following features:* (*i*) *Hiring occurs only in the bottom job, where workers perform Activity 1;* (*ii*) *There is a well-defined career path: bottom workers stay in the bottom job or are promoted. The promotion rate is constant and given by* $d_2 N_2/\left(\left(1 - d_1\right) N_1\right)$. *Top workers perform Activity 2 and are never demoted. Workers are not fired unless they are caught shirking.* (*iii*) *The performance bonus in the top job is constant and independent of the firm's span. The performance bonus in the bottom job is also constant, and it is equal to zero if the span* N_1/N_2 *is below a threshold and is otherwise positive and increasing in the span.*

PROOF OF PROPOSITION 2. We prove Proposition 2 in two steps.

Step 1. We first introduce some notations, list the relevant constraints for the problem, and establish a lower bound for the firm's objective function. Let

$$c\left(h^t\right) = \begin{cases} c_1 & \text{if } h_t = 1 \\ c_2 & \text{if } h_t = 2 \end{cases},$$

$$d(h^t) = \begin{cases} d_1 & \text{if } h_t = 1 \\ d_2 & \text{if } h_t = 2 \end{cases},$$

and

$$q\left(h^t\right) = \begin{cases} q_1 & \text{if } h_t = 1 \\ q_2 & \text{if } h_t = 2 \end{cases}.$$

Note that for any compensation plan with wage $\widetilde{w}\left(h^t\right)$ and bonus $\widetilde{b}\left(h^t\right)$, the firm can choose an alternative compensation plan with wage $\underline{w}$ and bonus $\widetilde{w}\left(h^t\right)+\widetilde{b}\left(h^t\right)-\underline{w}$. This new compensation plan gives the worker the same compensation (following all histories) and weakly improves the worker's incentive. To simplify notation, below we set the wage $w\left(h^t\right)=\underline{w}$ following all histories.

For the promise-keeping constraint for the worker (following history h^t), we have

$$v\left(h^t\right) = \underline{w} - c\left(h^t\right) + b\left(h^t\right) + \delta\left(1-d(h^t)\right)\left(p_1\left(h^t\right)v\left(h^t,1\right)+p_2\left(h^t\right)v\left(h^t,2\right)\right), \qquad \text{(PK-}h^t\text{)}$$

where $v\left(h^t,j\right)$ $(j=1,2)$ denotes the worker's value when he is assigned to activity j following history h^t. To induce effort from the agent, we have

$$v\left(h^t\right) \geq \underline{w} + \left(1-q\left(h^t\right)\right)\left(b\left(h^t\right)+\delta\left(1-d(h^t)\right)\left(p_1\left(h^t\right)v\left(h^t,1\right)+p_2\left(h^t\right)v\left(h^t,2\right)\right)\right).$$

Using $(PK-h^t)$, we can rewrite this inequality above as

$$v\left(h^t\right) \geq \underline{w} + \frac{\left(1-q\left(h^t\right)\right)}{q\left(h^t\right)}c\left(h^t\right). \qquad \text{(IC-}h^t\text{)}$$

This gives the incentive constraint $\left(IC-h^t\right)$. Notice that (by the definition of q and c) the right hand side of the inequality depends on the history h^t only through h_t,the worker's activity in period t.

Next, to describe the flow constraint, define $L\left(h^t\right)$ as the mass of workers with history h^t. The flow constraint following history h^t can then be written as

$$\begin{aligned} L\left(h^t,1\right) &= \left(1-d(h^t)\right)p_1\left(h^t\right)L\left(h^t\right); && \text{(FL-}h^t-1\text{)} \\ L\left(h^t,2\right) &= \left(1-d(h^t)\right)p_2\left(h^t\right)L\left(h^t\right). && \text{(FL-}h^t-2\text{)} \end{aligned}$$

In addition, there are two aggregate flow constraints:

$$\begin{aligned} \sum_{h^t|h_t=1} L\left(h^t\right) &= N_1; && \text{(FL-1)} \\ \sum_{h^t|h_t=2} L\left(h^t\right) &= N_2. && \text{(FL-2)} \end{aligned}$$

Given these constraints and a given organizational structure, the firm chooses (nonnegative) bonuses $\left(b\left(h^t\right)\right)$ and assignment rule $\left(p_1\left(h^t\right), p_2\left(h^t\right)\right)$ to minimize the total wage payment

$$\sum_{h^t} L\left(h^t\right)\left(\underline{w}+b\left(h^t\right)\right).$$

Notice that $(FL-1)$ and $(FL-2)$ imply that $\sum_{h^t} L\left(h^t\right)\underline{w} = (N_1+N_2)\,\underline{w}$. We can therefore rewrite the firm's objective as to minimize the total bonus

$$\sum_{h^t} L\left(h^t\right) b\left(h^t\right).$$

We now establish a lower bound for the objective function. Multiplying $L\left(h^t\right)$ to both sides of the promise-keeping constraint $\left(PK-h^t\right)$ and rearrange, we have

$$\begin{aligned}
& L\left(h^t\right) b\left(h^t\right) \\
= \; & L\left(h^t\right) v\left(h^t\right) + L\left(h^t\right) c\left(h^t\right) - \delta\left(1-d(h^t)\right) L\left(h^t\right)\left(p_1\left(h^t\right) v\left(h^t,1\right) + p_2\left(h^t\right) v\left(h^t,2\right)\right) - L(h^t)\underline{w} \\
= \; & L\left(h^t\right) v\left(h^t\right) + L\left(h^t\right) c\left(h^t\right) - \delta L\left(h^t,1\right) v\left(h^t,1\right) - \delta L\left(h^t,2\right) v\left(h^t,2\right) - L(h^t)\underline{w},
\end{aligned}$$

where the second equality follows from the flow constraints $\left(FL-h^t-1\right)$ and $\left(FL-h^t-2\right)$.

This implies that

$$\begin{aligned}
& \sum_{h^t} L\left(h^t\right) b\left(h^t\right) \\
= \; & \sum_{h^t}\left(L\left(h^t\right) v\left(h^t\right) + L\left(h^t\right) c\left(h^t\right) - \delta L\left(h^t,1\right) v\left(h^t,1\right) - \delta L\left(h^t,2\right) v\left(h^t,2\right)\right) - \sum_{h^t} L\left(h^t\right)\underline{w} \\
= \; & \delta\left(L(1)\,v(1) + L(2)\,v(2)\right) + (1-\delta)\sum_{h^t} L\left(h^t\right) v\left(h^t\right) + \sum_{h^t} L\left(h^t\right) c\left(h^t\right) - \sum_{h^t} L\left(h^t\right)\underline{w} \\
= \; & \delta\left(L(1)\,v(1) + L(2)\,v(2)\right) + (1-\delta)\sum_{h^t} L\left(h^t\right) v\left(h^t\right) + N_1 c_1 + N_2 c_2 - N_1\underline{w} - N_2\underline{w},
\end{aligned}$$

where the second equality follows because

$$\sum_{h^t} L\left(h^t\right) v\left(h^t\right) = L(1)v(1) + \sum_{h^t} L\left(h^t,1\right) v\left(h^t,1\right) + L(2)v(2) + \sum_{h^t} L\left(h^t,2\right) v\left(h^t,2\right)$$

and the last equality follows because

$$\sum_{h^t} L\left(h^t\right) c\left(h^t\right) = \sum_{h^t|h_t=1} L\left(h^t\right) c\left(h^t\right) + \sum_{h^t|h_t=2} L\left(h^t\right) c\left(h^t\right).$$

To establish a lower bound for $\sum_{h^t} L\left(h^t\right) b\left(h^t\right)$, below we separately provide a lower bound for $\sum_{h^t} L\left(h^t\right) v\left(h^t\right)$ and for $L(1)\,v(1) + L(2)\,v(2)$. Denote the incentive rent on activity i by

$$r_i \equiv \frac{1-q_i}{q_i} c_i,$$

for $i=1,2$. It follows from the IC constraint that

$$\sum_{h^t} L\left(h^t\right) v\left(h^t\right) = \sum_{h^t|h_t=1} L\left(h^t\right) v\left(h^t\right) + \sum_{h^t|h_t=2} L\left(h^t\right) v\left(h^t\right) \geq N_1(\underline{w}+r_1) + N_2(\underline{w}+r_2).$$

Next, notice that the mass of new workers $(L(1)+L(2))$ must exceed the the mass of workers who leave voluntarily, i.e.,

$$L(1) + L(2) \geq d_1 N_1 + d_2 N_2.$$

Given that $r_2 \geq r_1$, we then have

$$L(1)\, v(1) + L(2)\, v(2) \geq (d_1 N_1 + d_2 N_2)(r_1 + \underline{w}).$$

Combining these two lower bounds, we now have that

$$\begin{aligned}
& \sum_{h^t} L\left(h^t\right) b\left(h^t\right) \\
= \;& \delta\left(L(1)\, v(1) + L(2)\, v(2)\right) + (1-\delta) \sum_{h^t} L\left(h^t\right) v\left(h^t\right) + N_1 c_1 + N_2 c_2 - N_1 \underline{w} - N_2 \underline{w} \\
\geq \;& \delta(d_1 N_1 + d_2 N_2)(r_1 + \underline{w}) + (1-\delta)\left(N_1 r_1 + N_2 r_2 + (N_1 + N_2)\underline{w}\right) \\
& + N_1 c_1 + N_2 c_2 - N_1 \underline{w} - N_2 \underline{w}.
\end{aligned}$$

Thus, if we can choose a set of feasible $\left\{b\left(h^t\right), p_1\left(h^t\right), p_2\left(h^t\right)\right\}_{t=1}^{\infty}$ such that $\sum_{h^t} L\left(h^t\right) b\left(h^t\right)$ reaches the lower bound above, this contract is optimal.

Step 2. We show in this step that the following stationary contract is optimal, and demonstrate in the end that this set of contracts lead to the optimal personnel policy in Proposition 2.

In particular, for h^t with $h_t = 2$, let

$$b\left(h^t\right) = b_2,\ p_1\left(h^t\right) = 0,\ p_2\left(h^t\right) = 1,$$

where b_2 is given by

$$b_2 = c_2 + r_2 - \delta\left(1 - d_2\right)\left(r_2 + \underline{w}\right).$$

Notice that by Assumption 3, we have $b_2 \geq 0$.

For h^t with $h_t = 1$, let

$$b\left(h^t\right) = \max\{0, b_1\},\ p_1\left(h^t\right) = 1 - p,\ p_2\left(h^t\right) = p,$$

where

$$p = \frac{d_2 N_2}{\left(1 - d_1\right) N_1},$$

and

$$b_1 = c_1 + r_1 - \delta\left(1 - d_1\right)\left(p r_2 + (1-p)\, r_1 + \underline{w}\right).$$

Finally, let $L(1) = (1 - d_1) N_1 + (1 - d_2) N_2$ and $L(2) = 0$, and this completes the description of the contract (and the associated personnel policies). It is straightforward to check that the contract satisfies all the constraints and is therefore feasible. In addition, this contract gives rise to the properties (i) to (iii). It remains to show that the contract is optimal. There are two cases to consider.

Case 1: $b_1 \geq 0$. In this case, we have

$$\begin{aligned}
& \sum_{h^t|h_t=1} L\left(h^t\right) b\left(h^t\right) \\
= \; & N_1 b_1 \\
= \; & N_1 r_1 + N_1 c_1 - \delta\left(1-d_1\right) N_1\left(p r_2 + (1-p) r_1 + \underline{w}\right) \\
= \; & N_1 r_1 + N_1 c_1 - \delta d_2 N_2\left(r_2 - r_1\right) - \delta\left(1-d_1\right) N_1 r_1 - \delta\left(1-d_1\right) N_1 \underline{w}.
\end{aligned}$$

and

$$\begin{aligned}
& \sum_{h^t|h_t=2} L\left(h^t\right) b\left(h^t\right) \\
= \; & N_2 b_2 \\
= \; & N_2 r_2 + N_2 c_2 - \delta\left(1-d_2\right) N_2 r_2 - \delta\left(1-d_2\right) N_2 \underline{w}.
\end{aligned}$$

Therefore, it follows

$$\begin{aligned}
& \sum_{h^t} L\left(h^t\right) b\left(h^t\right) \\
= \; & N_1 b_1 + N_2 b_2 \\
= \; & N_1 r_1 + N_1 c_1 - \delta d_2 N_2\left(r_2 - r_1\right) - \delta\left(1-d_1\right) N_1 r_1 - \delta\left(1-d_1\right) N_1 \underline{w} \\
& + N_2 r_2 + N_2 c_2 - \delta\left(1-d_2\right) N_2 r_2 - \delta\left(1-d_2\right) N_2 \underline{w}. \\
= \; & \delta(d_1 N_1 + d_2 N_2)(r_1 + \underline{w}) + (1-\delta)\left(N_1 r_1 + N_2 r_2 + (N_1 + N_2)\underline{w}\right) \\
& + N_1 c_1 + N_2 c_2 - N_1 \underline{w} - N_2 \underline{w}.
\end{aligned}$$

which reaches the lower bound of $\sum_{h^t} L\left(h^t\right) b\left(h^t\right)$.

Case 2: $b_1 < 0$. In this case, we show below that the bonus amount in the contract above reaches a lower bound for the bonus amount for a relaxed problem. The relaxed problem has the same objective function and the constraints as the original problem, except that the firm does not consider the worker's incentive constraints when he is assigned to activity 1 ($h_t = 1$). As a result, the total bonus associated with optimal contract for the relaxed problem must be weakly lower. It follows that if the bonus with the contract reaches the lower bound of bonus for the relaxed problem, the contract must be optimal.

Now consider a contract with bonus $\left(b\left(h^t\right)\right)$ and assignment rule $\left(p_1\left(h^t\right), p_2\left(h^t\right)\right)$ for the relaxed problem, and let $v\left(h^t\right)$ be the associated value function for the worker. Notice that for $h_t = 2$, the flow constraints and the promise-keeping constraints imply that

$$L\left(h^t\right) v(h^t) = L\left(h^t\right)\left(\underline{w} + b\left(h^t\right) - c_2\right) + \delta\left(L\left(h^t, 1\right) v\left(h^t, 1\right) + L\left(h^t, 2\right) v\left(h^t, 2\right)\right).$$

Next, we establish the lower bounds in two steps. First, we consider a subset of contracts with a particular property and show that the optimal contracts satisfy this property. We then provide a lower bound for the wage bills for this subset of contracts. Since this subset of contracts contains the optimal contracts, a lower bound of the wage bill for this subset of contracts is also the lower bound for all contracts.

Now, consider the class of contracts where $v\left(h^t, 1\right) \leq \underline{w} + r_2$ and $v\left(h^t, 2\right) = \underline{w} + r_2$ for all

$h^t \neq \emptyset$. We show that this class of constraints contains the optimal contracts. To see this, notice that following history h^t, we have

$$L\left(h^t\right) v(h^t) = \sum_{h^\tau | h^t,} \delta^{\tau - t} L\left(h^\tau | h^t\right) \left(\underline{w} + b\left(h^\tau | h^t\right) - c\left(h^\tau | h^t\right)\right),$$

where $h^\tau | h^t$ are histories consistent with h^t. Now consider the following perturbation on $b\left(h^\tau | h^t\right)$ and $b\left(h^t\right)$. By reducing $b\left(h^\tau | h^t\right)$ to $b\left(h^\tau | h^t\right) - \varepsilon$ and increasing $b\left(h^t\right)$ to $b\left(h^t\right) + \frac{\delta^{\tau-t} L\left(h^\tau | h^t\right)}{L(h^t)} \varepsilon$, the equality above continues to hold. The total bonus the firm pays, however, changes by

$$L\left(h^t\right) \frac{\delta^{\tau-t} L\left(h^\tau | h^t\right)}{L\left(h^t\right)} \varepsilon - L\left(h^\tau | h^t\right) \varepsilon = -\left(1 - \delta^{\tau - t}\right) L\left(h^\tau | h^t\right) \varepsilon \leq 0.$$

In other words, if this perturbation is feasible, the contract suboptimal.

Now suppose that $v\left(h^t, 2\right) > \underline{w} + r_2$ so that the incentive constraint is slack. It follows that if $b\left(h^t, 2\right) > 0$, the firm can use the type of perturbation above on $b\left(h^t, 2\right)$ and $b\left(h^t\right)$ since the incentive constraint on $\left(h^t, 2\right)$ is slack. If $b\left(h^t, 2\right) = 0$, the promise-keeping condition on $v\left(h^t, 2\right)$ (since it exceeds $\underline{w} + r_2$) implies that there either exists some $b\left(h^\tau | h^t, 2\right) > 0$ with $h_\tau = 1$ or some $b\left(h^{\tau'} | h^t, 2\right) > 0$ with $h_{\tau'} = 2$ and $v\left(h^{\tau'} | h^t, 2\right) > \underline{w} + r_2$. Otherwise, it contradicts the assumption that $v\left(h^t, 2\right) > \underline{w} + r_2$. In the former case, the firm can perform a feasible perturbation on $b\left(h^\tau | h^t, 2\right)$ and $b\left(h^t\right)$ as above. This is feasible because there is no incentive constraints on activity 1. In the later case, the firm can perform a feasible perturbation on $b\left(h^{\tau'} | h^t, 2\right)$ and $b\left(h^t\right)$. This reduces the firm's total bonus payout, and therefore, contradicts the optimality of the original contract.

Similarly, consider the case with $v\left(h^t, 1\right) > \underline{w} + r_2$. If $b\left(h^t, 1\right) > 0$, then the firm can use the perturbation above on $b\left(h^t, 1\right)$ and $b\left(h^t\right)$. If $b\left(h^t, 1\right) = 0$, Assumption 3 then implies that there either exists some $b\left(h^\tau | h^t, 1\right) > 0$ with $h_\tau = 1$ or some $b\left(h^{\tau'} | h^t, 1\right) > 0$ with $h_{\tau'} = 2$ and $v\left(h^{\tau'} | h^t, 1\right) > \underline{w} + r_2$. By performing the perturbation on $b\left(h^t\right)$ and $b\left(h^\tau | h^t, 1\right)$ (or on $b\left(h^t\right)$ and $b\left(h^{\tau'} | h^t, 1\right)$), the firm can again reduce its bonus payout, and therefore, contradicts the optimality of the original contract. This shows that under the optimal contracts belong to the set with $v\left(h^t, 1\right) \leq \underline{w} + r_2$ and $v\left(h^t, 2\right) = \underline{w} + r_2$.

Next, we establish a lower bound of wage bills for this subset of contracts. Notice that

$$\begin{aligned}
& \sum_{h^t} L\left(h^t\right) b\left(h^t\right) \\
\geq \quad & \sum_{h^t | h_t = 2} L\left(h^t\right) b\left(h^t\right) \\
= \quad & L\left(h^t\right) v\left(h^t\right) + L\left(h^t\right) c_2 - L\left(h^t\right) \underline{w} - \delta\left(1 - d_2\right) L\left(h^t\right) \left(p_1\left(h^t\right) v\left(h^t, 1\right) + p_2\left(h^t\right) v\left(h^t, 2\right)\right) \\
\geq \quad & L\left(h^t\right) \left(\underline{w} + r_2\right) + L\left(h^t\right) c_2 - L\left(h^t\right) \underline{w} - \delta\left(1 - d_2\right) L\left(h^t\right) \left(\underline{w} + r_2\right) \\
= \quad & N_2(r_2 + c_2) - \delta\left(1 - d_2\right) N_2(\underline{w} + r_2),
\end{aligned}$$

where the last inequality follows because $v\left(h^t,1\right)\leq \underline{w}+r_2$ and $v\left(h^t,2\right)=\underline{w}+r_2$.

In addition, one can check that the contract constructed reach this lower bound. The contract is therefore optimal.■

References

[1] Agrawal, Anup, Charles Knoeber, and Theofanis Tsoulouhas. 2006. Are Outsiders Handicapped in CEO Successions? *Journal of Corporate Finance*, 12(3): 619-644.

[2] Akerlof, George and Lawrence Katz. 1989. Workers' Trust Funds and the Logic of Wage Profiles. *Quarterly Journal of Economics*, 104(3): 525-536.

[3] Auriol, Emmanuelle, Guido Friebel, and Frauke von Bieberstein. 2013. The Firm as the Locus of Social Comparisons: Internal Labor Markets versus Up-or-Out. *Working Paper*.

[4] Baker, George, Michael Jensen, and Kevin Murphy. 1988. Compensation and Incentives: Practice vs. Theory. *Journal of Finance*, 43(3): 593-616.

[5] Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994a. The Internal Economics of the Firm: Evidence from Personnel Data. *Quarterly Journal of Economics*, 109(4): 881-919.

[6] Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994b. The Wage Policy of a Firm. *Quarterly Journal of Economics*, 109(4): 921-955.

[7] Banks, James, Richard Blundell, Antoine Bozio, and Carl Emmerson. 2010. Releasing Jobs for the Young? Early Retirement and Youth Unemployment in the United Kingdom. In *Social Security Programs and Retirement Around the World: The Relationship to Youth Employment*, eds. Jonathan Gruber and David Wise. Chicago: University of Chicago Press.

[8] Baron, James and William Bielby. 1986. The Proliferation of Job Titles in Organizations. *Administrative Science Quarterly*, 31(4): 561-586.

[9] Barron, John, Dan Black, and Mark Loewenstein. 1987. Employer Size: The Implications for Search, Training, Capital Investment, Starting Wages, and Wage Growth. *Journal of Labor Economics*, 5(1): 76-89.

[10] Bartholomew, David. 1973. *Stochastic Models for Social Processes*. New York: Wiley.

[11] Becker, Gary. 1975. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. University of Chicago Press.

[12] Belzil, Christian and Michael Bognanno. 2008. Promotions, Demotions, Halo Effects, and the Earnings Dynamics of American Executives. *Journal of Labor Economics*, 26(2): 287-310.

[13] Bernhardt, Dan. 1995. Strategic Promotion and Compensation. *Review of Economic Studies*, 62(2): 315-339.

[14] Blau, Peter and Richard Schoenherr. 1971. *The Structure of Organizations*. New York: Basic Books.

[15] Board, Simon. 2011. Relational Contracts and the Value of Loyalty. *American Economic Review*, 101(7): 3349-3367.

[16] Bolton, Patrick and Mathias Dewatripont. 2004. *Contract Theory*. MIT Press.

[17] Bond, Timothy. 2014. Internal Labor Markets in Equilibrium. *Working Paper*.

[18] Brown, Charles and James Medoff. 1989. The Employer Size Wage Effect. *Journal of Political Economy*, 97(5): 1027-1059.

[19] Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. Forthcoming. The Anatomy of French Production Hierarchies. *Journal of Political Economy*.

[20] Calvo, Guillermo and Stanislaw Wellisz. 1978. Supervision, Loss of Control, and the Optimum Size of the Firm. *Journal of Political Economy*, 86(5): 943-952.

[21] Camara, Odilon and Dan Bernhardt. 2009. The Dynamics of Promotions, Quits, and Layoffs. *Working Paper*.

[22] Cappelli, Peter. 2008. *Talent on Demand*. Harvard Business Press.

[23] Carmichael, Lorne. 1985. Can Unemployment Be Involuntary?: Comment. *American Economic Review*, 75(5): 1213-1214.

[24] Chan, William. 1996. External Recruitment versus Internal Promotion. *Journal of Labor Economics*, 14(4): 555-570.

[25] Colombo, Massimo and Marco Delmastro. 1999. Some Stylized Facts on Organization and its Evolution. *Journal of Economic Behavior and Organization*, 40(3): 255-275.

[26] Dalton, Dan and Idalene Kesner. 1983. Inside/Outside Succession and Organizational Size: The Pragmatics of Executive Replacement. *Academy of Management Journal*, 26(4): 736-742.

[27] Demougin, Dominique and Aloysius Siow. 1994. Careers in Ongoing Hierarchies. *American Economic Review*, 84(5): 1261-1277.

[28] DeVaro, Jed and Hodaka Morita. 2013. Internal Promotion and External Recruitment: A Theoretical and Empirical Analysis. *Journal of Labor Economics*, 31(2): 227-269.

[29] Doeringer, Peter and Michael Piore. 1971. *Internal Labor Markets and Manpower Analysis*. Heath Lexington Books.

[30] Dohmen, Thomas, Ben Kriechel, and Gerard Pfann. 2004. Monkey Bars and Ladders: The Importance of Lateral and Vertical Job Mobility in Internal Labor Market Careers. *Journal of Population Economics*, 17: 193-228.

[31] Fong, Yuk-Fai and Jin Li. 2014. Relational Contracts, Efficiency Wages, and Employment Dynamics. *Working Paper*.

[32] Gabaix, Xavier and Augustin Landier. 2008. Why Has CEO Pay Increased So Much? *Quarterly Journal of Economics*, 123(1): 49-100.

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

[34] Garicano, Luis and Thomas Hubbard. 2009. Earnings Inequality and Coordination Costs: Evidence from U.S. Law Firms. *Working Paper.*

[35] Gibbons, Robert. 1997. Incentives and Careers in Organizations. In *Advances in Econometrics: Theory and Applications*, eds. David Kreps and Kenneth Wallis, Cambridge University Press.

[36] Gibbons, Robert and Michael Waldman. 1999a. A Theory of Wage and Promotion Dynamics inside Firms. *Quarterly Journal of Economics*, 114(4): 1321-1358.

[37] Gibbons, Robert and Michael Waldman. 1999b. Careers in Organizations: Theory and Evidence. In *Handbook of Labor Economics, vol 3*, eds. Orley Ashenfelter and David Card, North Holland.

[38] Gibbs, Michael. 1995. Incentive Compensation in a Corporate Hierarchy. *Journal of Accounting and Economics*, 19(2): 247-277.

[39] Gruber, Jonathan and David Wise. 2010. *Social Security Programs and Retirement around the World: The Relationship to Youth Unemployment.* University of Chicago Press.

[40] Guadalupe, Maria and Julie Wulf. 2010. The Flattening Firm and Product Market: The Effect of Trade Liberalization on Corporate Hierarchies. *American Economic Journal: Applied Economics*, 2(4): 105-127.

[41] Harris, Milton and Bengt Holmstrom. 1982. A Theory of Wage Dynamics. *Review of Economic Studies*, 49(3): 315-333.

[42] Huson, Mark, Paul Malatesta, and Robbert Parrino. 2004. Managerial Succession and Firm Performance. *Journal of Financial Economics*, 74(2): 237-275.

[43] Imai, Kenichi and Itami Hiroyuki. 1988. Allocation of Labor and Capital in Japan and the United States. In *Inside the Japanese System*, eds. Daniel Okimoto and Thomas Rohlen, Stanford University Press: 112-117.

[44] Jensen, Michael. 1986a. Agency Costs of Free Cash Flow, Corporate Finance and Takeovers." *American Economic Review*, 76(2): 323-329.

[45] Jensen, Michael. 1986b. The Takeover Controversy: Analysis and Evidence. *Midland Corporate Finance Journal*, 4(2): 6-32.

[46] Kasarda, John. 1974. The Structural Implications of Social System Size: A Three-Level Analysis. *American Sociological Review*, 39(1): 19-28.

[47] Keyfitz, Nathan. 1973. Individual Mobility in a Stationary Society. *Population Studies: A Journal of Demography*, 27(2): 335-52.

[48] Krakel, Matthias and Anja Schottner. 2012. Internal Labor Markets and Worker Rents. *Journal of Economic Behavior and Organization*, 84(2): 491-509.

[49] LAUTERBACH, BENI AND JACOB WEISBERG. 1994. Top Management Successions: The Choice Between Internal and External Sources. *International Journal of Human Resource Management*, 5(1): 51-65.

[50] LAUTERBACH, BENI, JOSEPH VU, AND JACOB WEISBERG. 1999. Internal vs. External Successions and Their Effect on Firm Performance. *Human Relations*, 52(12): 1485-1504.

[51] LAZEAR, EDWARD. 1979. Why is there Mandatory Retirement? *Journal of Political Economy*, 87(6): 1261-1284.

[52] LAZEAR, EDWARD. 1992. "The Job as a Concept," in Wiliam J. Bruns, Jr., ed., Performance measurement, evaluation, and incentives. Boston, MA: Harvard Business School Press, , pp. 183-215.

[53] LAZEAR, EDWARD. 1999. Personnel Economics: Past Lessons and Future Directions. NBER Working Paper w6957.

[54] LAZEAR, EDWARD AND ROBERT MOORE. 1984. Incentives, Productivity, and Labor Contracts. *Quarterly Journal of Economics*, 99(2): 275-296.

[55] LAZEAR, EDWARD AND PAUL OYER. 2013. Personnel Economics. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.

[56] LAZEAR, EDWARD AND SHERWIN ROSEN. 1981. Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, 89(5): 841-864.

[57] LAZEAR, EDWARD, KATHRYN SHAW, AND CHRISTOPHER STANTON. 2013. Making Do With Less: Working Harder During Recessions. *Working Paper*.

[58] LIN, MING-JEN. 2005. Opening the Black Box: The Internal Labor Markets of Company X. *Industrial Relations*, 44(4): 659-706.

[59] MACLEOD, BENTLEY AND JAMES MALCOMSON. 1988. Reputation and Hierarchy in Dynamic Models of Employment. *Journal of Political Economy*, 96(4): 832-854.

[60] MALCOMSON, JAMES. 1984. Work Incentives, Hierarchy, and Internal Labor Markets. *Journal of Political Economy*, 92(3): 486-507.

[61] MCCUE, KRISTIN. 1996. Promotions and Wage Growth. *Journal of Labor Economics*, 14(2): 175-209.

[62] MEDOFF, JAMES AND KATHARINE ABRAHAM. 1980. Experience, Performance, and Earnings. *Quarterly Journal of Economics*, 95(4): 703-736.

[63] MEDOFF, JAMES AND KATHARINE ABRAHAM. 1981. Are Those Paid More Really More Productive: The Case of Experience. *Journal of Human Resources*, 16(2): 186-216.

[64] MILGROM, PAUL AND JOHN ROBERTS. 1992. *Economics, Organization and Management*. Prentice Hall.

[65] Mookherjee, Dilip. 2013. Incentives in Hierarchies. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.

[66] Moore, Henry. 1911. *Laws of Wages*. New York: Macmillan.

[67] Morris, Stuart. 1973. Stalled Professionalism: The Recruitment of Railway Officials in the United States, 1885-1940. *Business History Review*, 47(3): 317-334.

[68] Murphy, Kevin. 1985. Corporate Performance and Managerial Remuneration: An Empirical Analysis. *Journal of Accounting and Economics*, 7(1): 11-42.

[69] Oi, Walter and Todd Idson. 1999. Firm Size and Wages. In *Handbook of Labor Economics*, eds. Orley Ashenfelter and David Card. Elsevier Science.

[70] Oyer, Paul. 2007. Is There an Insider Advantage in Getting Tenure? *American Economic Review*, 97(2): 501-505.

[71] Parrino, Robert. 1997. CEO Turnover and Outside Succession: A Cross-Sectional Analysis. *Journal of Financial Economics*, 46(2): 165-197.

[72] Prendergast, Canice. 1993. The Role of Promotion in Inducing Specific Human Capital Acquisition. *Quarterly Journal of Economics*, 108(2): 523-534.

[73] Qian, Yingyi. 1994. Incentives and Loss of Control in an Optimal Hierarchy. *Review of Economic Studies*, 61(3): 527-544.

[74] Rebitzer, James and Lowell Taylor. 1995. Efficiency Wages and Employment Rents: The Employer-Size Wage Effect in the Job Market for Lawyers. *Journal of Labor Economics*, 13(4): 678-708.

[75] Rosen, Sherwin. 1986. Prizes and Incentives in Elimination Tournaments. *American Economic Review*, 76(4): 701-715.

[76] Rosenbaum, James. 1984. *Career Mobility in a Corporate Hierarchy*. New York: Academic.

[77] Rushing, William. 1966. Organizational Size and Administration: The Problems of Causal Homogeneity and a Heterogeneous Category. *Pacific Sociological Review*, 9(2): 100-108.

[78] Seltzer, Andre and David Merrett. 2000. Personnel Policies at the Union Bank of Australia: Evidence from the 1888-1900 Entry Cohorts. *Journal of Labor Economics*, 18(4): 573-613.

[79] Shapiro, Carl and Joseph Stiglitz. 1984. Equilibrium Unemployment as a Worker Discipline Device. *American Economic Review*, 74(3): 433-444

[80] Simon, Herbert. 1951. The Analysis of Promotional Opportunities. *Personnel*, 27: 282-285.

[81] Stewman, Shelby. 1986. Demographic Models of Internal Labor Markets. *Administrative Science Quarterly*, 31(2): 212-47.

[82] Stewman, Shelby. 1988. Organizational Demography. *Annual Review of Sociology* 14:173-202.

[83] Stewman, Shelby. and Suresh Konda. 1983. Careers and Organizational Labor Markets: Demographic Models of Organizational Behavior. *American Journal of Sociology*, 88(4): 637-85.

[84] Tervio, Marko. 2008. The Difference That CEOs Make: An Assignment Model Approach. *American Economic Review*, 98(3): 642-668.

[85] Treble, John, Edwin van Gameren, Sarah Bridges, and Tim Barmby. 2001. The Internal Economics of the Firm: Further Evidence from Personnel Data. *Journal of Labor Economics*, 8(5): 531-552.

[86] U.S. Department of Labor. 1981. Abolishing Mandatory Retirement. Washington, U.S.: Government Printing Office.

[87] Waldman, Michael. 1984. Job Assignments, Signalling, and Efficiency. *RAND Journal of Economics*, 15(2): 255-267.

[88] Waldman, Michael. 2003. Ex Ante versus Ex Post Optimal Promotion Rules: The Case of Internal Promotion. *Economic Inquiry*, 41(1): 27-41.

[89] Waldman, Michael. 2013. Classic Promotion Tournaments versus Market-Based Tournaments. *International Journal of Industrial Organization*, 31(3): 198-210.

[90] Williamson, Oliver. 1967. *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm*. Chicago, Illinois: Markham Publishing Company.

[91] White, Harrison. 1970. *Chains of Opportunity: System Models of Mobility in Organizations.* Cambridge, MA: Harvard University Press.

[92] Zabojnik, Jan and Dan Bernhardt. 2001. Corporate Tournaments, Human Capital Acquisition, and the Firm Size-Wage Relation. *Review of Economic Studies*, 68(3): 693-716.