

Managing Careers in Organizations*

Rongzhu Ke

Jin Li

Michael Powell

January 17, 2016

Abstract

Firms' organizational structures impose constraints on their ability to use promotion-based incentives. We develop a framework for identifying these constraints and exploring their consequences. We show that firms manage workers careers by choosing personnel policies that resemble an internal labor market. Firms may adopt forced-turnover policies to keep lines of advancement open, and 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.

Keywords: internal labor markets, promotions, dynamic incentives

JEL classifications: D86, J41, M51

*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 Urgan for excellent research assistance. We are grateful to Daron Acemoglu, Ricardo Alonso, Pierre Azoulay, Dan Barron, Tim Bond, Jennifer Brown, Jimmy Chan, Jed DeVaro, 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 Zabojsnik, 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 so as to cope with these constraints have received little attention.¹ In this paper, we develop a framework in which promotion opportunities motivate workers, but firms face a “budget constraint” on promotion opportunities. Firms manage workers’ careers subject to this constraint, and they can alter their organizational structure in order to relax this constraint. Our model 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. For promotions to arise optimally, we incorporate two elements. First, production requires workers to perform multiple activities. Second, motivating workers performing a given activity requires providing them with rents, for example, because monetary transfers are constrained. Using familiar reasoning from dynamic moral-hazard models, we show that these rents should be backloaded in a worker’s career. A firm that is unable to backload rents perfectly within an activity will optimally backload them across activities: promotions arise as a way of managing the rents provided to workers.

We then turn to the heart of our problem and highlight a 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

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

their production to 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. We show that firms balance this trade-off optimally by choosing their personnel policies and the number of workers performing each activity. 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 harder to do or harder to monitor than the other (the low-rent activity). For example, the high-rent activity may involve solving difficult, non-routine problems or overseeing others. The firm’s output and therefore its revenues depend on how many workers perform each activity.

To maximize its steady-state profits, 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 and Implications. Optimal personnel policies reflect a tension between worker motivation and productive efficiency and exhibit two sets of features that 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. 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. 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 starting 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. Workers' wages therefore increase upon promotion.

The second set of features of optimal personnel policies arises 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. Forced-turnover policies, however, 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, firms might bias hiring into top jobs towards away from more-productive outsiders, precisely because hiring externally into the top job limits the firm's ability to use promotions for incentive purposes.

A key feature of our analysis is that optimal personnel policies depend on firm-level characteristics, so we can go beyond the claim that certain policies *might* arise, and we can say *when* they might arise. That is, our model describes how firm-level characteristics, such as its organizational span—the ratio of the number of positions at the bottom to the number of positions at the top—impact the types of personnel policies firms put in place. Firms with larger spans will naturally have more limited promotion opportunities. As such, they are more likely to put in place forced-turnover policies and practice stronger insider bias in hiring. These features imply that factors external to a given worker can shape a worker's career, even if workers are homogeneous.

In addition to choosing personnel policies, firms 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 is optimally lower than would be productively efficient for the wages it pays.

Extensions. A key result in our main model is that firms offer workers rents in the top job in

order to motivate them. In our main model, these rents arise because we assume wages are tied to the activities workers perform. The assumption that wages are tied to activities is stronger than we need in order to ensure that the firm has to provide workers with rents in the top job. In one extension, we consider a setting in which the firm is able to offer a general class of contracts involving history-dependent activity-assignment, wage, and bonus schemes. Optimal contracts in this setting again resemble an internal labor market that is designed to reuse incentive rents optimally. We also extend our model to consider industry equilibrium in a setting where many firms compete with each other in the product and labor markets, and they each choose optimal personnel policies to motivate their workers. We show that the features of optimal personnel policies in the main model arise in a setting in which all workers and firms earn zero ex-ante rents. Finally, we discuss how our main model can incorporate worker heterogeneity and capture a number of empirical regularities concerning wage and promotion dynamics that cannot be explained in our baseline model in which workers are homogeneous and wages are tied to activities.

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). As in these papers, ex ante promises of wage increases upon promotion serve as an incentive device. Unlike in Lazear and Rosen (1981), however, the firm also finds it optimal, ex post, to offer high wages to promoted workers. 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.

Since high wages in the top job serve multiple roles, our paper is also related to the literature on market-based promotion tournaments (Waldman, 1984; Gibbs, 1995; Zabochnik and Bernhardt, 2001; Ghosh and Waldman, 2010; Waldman 2013) in which promotions serve as a positive signal about the worker’s ability, in turn increasing his outside option and therefore his wage within the firm. It is also related to the literature on influence activities (Milgrom, 1988; Milgrom and Roberts, 1988) in which promoted workers are paid higher wages in order to reduce the probability that they take a more-appealing outside option. In contrast, in our paper, high wages in the top job serve the role of providing continuing incentives to workers in that job.

Our model focuses on how firms structure internal labor markets to provide incentives to workers (Malcomson, 1984; MacLeod and Malcomson, 1988; Prendergast, 1993; Zabochnik and Bernhardt,

2001; Waldman, 2003; Krakel and Schottner, 2012; Auriol, Friebe, and von Bieberstein, 2013). In contrast to these papers, factors of production in our model are flexible but subject to diminishing returns. This allows us to highlight a tension between incentive provision and productive efficiency, which generates systematic relationships between firms' 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 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; Fong and Li, 2015; Lazear, Shaw, and Stanton, Forthcoming). In our model, backloading pay occurs across activity assignments, and our setting is 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 structure 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.

2 The Model

A firm and a large mass of identical risk-neutral workers interact repeatedly. Time is discrete and denoted by t , and all players share discount factor $\delta \in (0, 1)$. In each period, the firm chooses its personnel policies, which we will describe below. We restrict attention to time-independent personnel policies 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 $F_{12} \geq 0$.

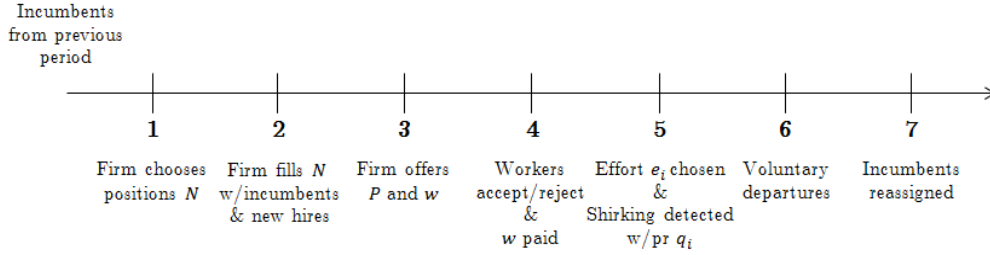


Figure 1: Timing of the stage game.

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.² 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 . Note that there are potentially two sources of turnover in the model, on the equilibrium path: voluntary departures and forced turnover. The total turnover rate for activity i is therefore $d_i + (1 - d_i)(1 - p_{i1} - p_{i2})$.

²In Section 7, we consider history-dependent compensation and promotion policies when workers are subject to a limited-liability constraint. We show that the optimal policy is stationary for workers who perform each activity.

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 which p_{12} and p_{21} are restricted to be 0. 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:

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

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

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

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, v_i . 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 requires that $v_i \geq R_i$, where

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

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 1. *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 $F_i(N_i^{pc}, N_j^{pc}) = w_i^{pc} > c_i$.*

Lemma 1 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

$$v_1 \geq 0; \quad (\text{IR-1})$$

$$v_2 \geq 0. \quad (\text{IR-2})$$

Incentive-Compatibility Constraints. Workers prefer to exert effort if their future payoffs exceed activity i 's incentive rents:

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

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

where R_i is the incentive rent for activity i .

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:

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

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

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 $p_{i1} + p_{i2} < 1$. Without loss of generality, we assume that $R_2 \geq R_1$.

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 choose 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 , so the assignment policy \mathbf{P} represents the firm's *promotion*, *demotion*, and *retention policies*. If $p_{i1} + p_{i2} = 1$, we say activity i has *full job security*: a worker performing activity i then 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:

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

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 2. *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.³

³An additional and natural reason for hiring directly into the bottom job that is beyond the scope of our main

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 3. *If 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 3 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 their incentive constraints are slack. The firm 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.

model is that prior experience performing activity 1 may be necessary for performing activity 2. Such a requirement would reinforce the result of Lemma 2.

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

LEMMA 4. *If 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, there is full job security at the bottom and forced turnover at the top.*

If there are insufficient separation rents, the personnel policies described in Lemma 3 do not provide enough motivation for bottom workers. To increase incentives for bottom workers, the firm could in principle pay higher wages at the bottom. Lemma 4 shows that doing so is never optimal—in an optimal personnel policy, bottom workers receive zero rents. The firm provides additional motivation entirely by increasing bottom workers' promotion prospects. To do so, the firm forces turnover among top workers 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 expand 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 are optimal, although demoting top workers is also optimal when the outside options of top workers are the same as those of bottom workers. Both forced-turnover policies and demotions 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 uniquely optimal, it has to be the case that bottom workers' outside options are greater than top workers' outside options.

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*

$$W(N_1, N_2) = c_1 N_1 + c_2 N_2 + (1 - \delta) N_2 R_2. \quad (\text{LC-1})$$

(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*

$$W(N_1, N_2) = c_1 N_1 + c_2 N_2 + (1 - \delta)((1 - d_1)N_1 R_1 + (1 - d_2)N_2 R_2). \quad (\text{LC-2})$$

Within each region 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 of adding a position in

activity i . In a Neoclassical model of labor supply in which there are no incentive problems, this coefficient would equal the wage for workers in activity i , which would equal 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 given by $(LC - 1)$ and when $N_1/N_2 \leq \sigma$, the labor-cost function is given by $(LC - 2)$. 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 charac-

terize a few general properties of the optimal organizational structure. We then draw connections between the firm's optimal production level and its optimal span, which will be relevant for our empirical discussion in Section 6.

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.

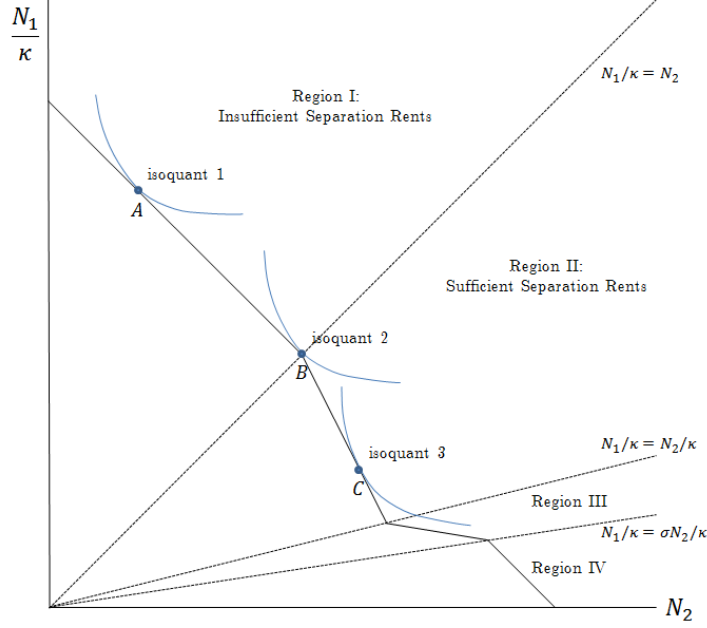


Figure 2: Producer-theory approach to a firm's cost-minimization problem. The isocost curve is piecewise linear with different coefficients on either side of the dotted boundaries. The isoquants represent different production technologies, with higher-numbered isoquants representing production technologies 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*

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

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

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

(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 back-loaded 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 $w_1^* = MRP_1^*$.

Next, we discuss how the production level (y) affects the firm’s optimal span $(N_1^*(y)/N_2^*(y))$. 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.

Given a labor-cost function, we can solve for the production-expansion path for general quasiconcave production functions using standard methods. In order to better link our model to the empirical findings we will discuss in Section 6, we will focus on production functions for which larger firms will tend to have larger spans. Assumption 2 provides a sufficient condition for a production function f to have an increasing and convex production-expansion path. Existing models of internal labor markets with endogenous production (Zabojnik and Bernhardt, 2001; DeVaro and Morita, 2003) restrict attention to production functions in which there is a fixed number of top positions in the firm, which 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.

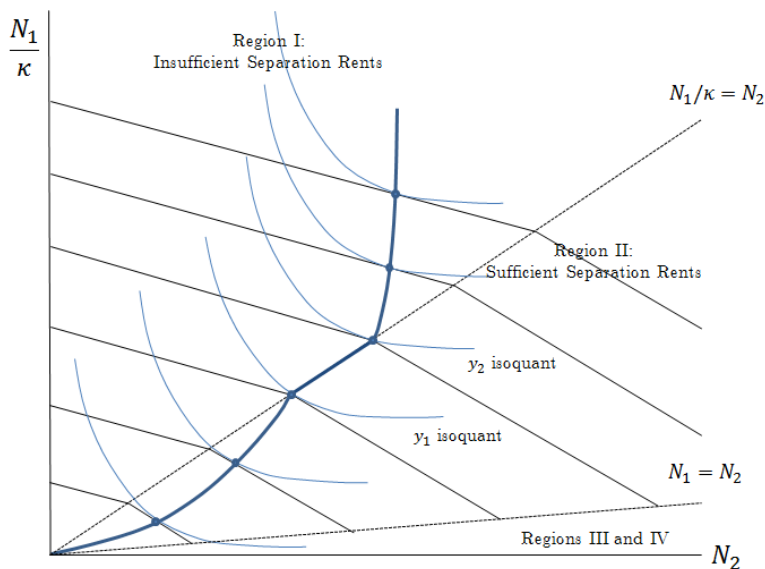


Figure 4: This figure plots a production-expansion path for a nonhomothetic production technology for which expansion favors activity 1. Firms with high production levels optimally operate in the insufficient-separation-rents region. Firms with low production levels optimally operate in the sufficient-separation-rents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

Figure 4 highlights the result, summarized in Corollary 4, 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 4. *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 5 summarizes key differences in the optimal personnel policies for these two firms.

COROLLARY 5. *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 5 to the empirical literature on how the careers of workers differ between small and large firms.

6 Empirical Relevance of Optimal Personnel Policies

In our model, optimal personnel policies exhibit a number of standard features that are often observed in empirical studies of internal labor markets (see, for example, Doeringer and Piore (1971)): workers tend to be hired into lower-level positions in the firm,⁴ there are promotions from within, and demotions are the exception rather than the norm. These broad patterns have motivated the development of many models of internal labor markets in which promotions serve many different roles. These existing models, however, pay limited attention to the constraints firms face in providing promotion opportunities, and our focus on these constraints highlights additional empirically relevant features of internal labor markets.

⁴The evidence for what Doeringer and Piore (1971) refers to as a "port of entry" is somewhat mixed (see, for example, Baker, Gibbs, and Holmstrom (1994)).

We first describe these features and the evidence supporting their prevalence, and then we assess the extent to which these features are consistent with existing theories of internal labor markets. A key feature of our analysis is that optimal personnel policies depend on firm-level characteristics in systematic ways. We conclude this section with a discussion of the evidence regarding the relationship between firm-level characteristics and personnel policies.

OBSERVATION 1. *There are wage jumps at promotion.*

In optimal personnel policies, the top job is more generously compensated than the bottom job. This wage differential serves the dual roles of motivating workers both at the top and at the bottom. Wage jumps at promotion are an empirical regularity in both managerial and non-managerial jobs. For evidence of wage jumps at promotion in managerial jobs, see Murphy (1985), Lazear (1992), Baker, Gibbs, and Holmstrom (1994a,b), and McCue (1996); for evidence of wage jumps at promotion in nonmanagerial jobs, see Dohmen (2004), Grund (2005), and Kwon (2006).

OBSERVATION 2. *Firms may exhibit an insider bias in hiring at the top.*

In the model, the firm has a strict preference for hiring only at the bottom. Even if workers are heterogeneous, optimal personnel policies would exhibit an insider bias in hiring at the top: an external hire into the top job would have to clear a higher hurdle than would internal candidates. Firms would bias hiring at the top for two reasons: firms must pay rents to new hires into the top job, but not to new hires into the bottom job. Further, hiring externally limits the firm's ability to use promotions for incentive purposes.

There is considerable evidence that firms exhibit an insider bias in hiring at the top. 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.

Importantly, it is worth noting that an insider bias in hiring at the top is not the same as a policy of promoting from within: there are many reasons why top positions might be filled by internal candidates, including firm-specific human capital and superior information about workers' ability. But it is less obvious why firms would want to promote from within even when they have the opportunity to hire a more-productive worker externally at the same wage.

OBSERVATION 3. *Firms may adopt forced-turnover policies.*

In our model, firms may adopt forced-turnover policies specifically to create promotion opportunities for workers. In the United States, prior to 1986 when this practice was outlawed, many firms

put in place mandatory-retirement policies. The leading explanation for mandatory-retirement policies is Lazear (1979), which argues that such policies can be 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 he would not willingly leave the firm without such a policy in place.

As Waldman (2003) points out, however, mandatory-retirement policies were often binding on the firm's side (as documented by Kittner (1977)). That is, they prohibited the firm from retaining the worker past a certain age. This feature is consistent with our model's rationale for mandatory-retirement policies: a binding commitment by the firm to dismiss older workers ensures promotion opportunities for younger workers. Our rationale is also consistent with stated reasons for adopting such policies. 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.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" 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.

Relationship to Existing Theories of Internal Labor Markets. There are many models of internal labor markets in which some of these types of personnel policies arise. Essentially all models of internal labor markets generate wage jumps at promotion, and they do so for many different reasons.⁵ See Gibbons and Waldman (1999b), Lazear (1999), Lazear and Oyer (2013), and Waldman (2013) for surveys.

In papers in which wage jumps at promotion serve specifically to motivate workers to exert

⁵Waldman (1984), Bernhardt (1995), Gibbons and Waldman (1999a), Ghosh and Waldman (2010), Bose and Lang (2013), and Bond (2015) provide symmetric- and asymmetric-learning-based explanations; Lazear and Rosen (1981), Malcomson (1984), Rosen (1986), MacLeod and Malcomson (1988), Milgrom and Roberts (1988), Prendergast (1993), Chan (1996), Zabochnik and Bernhardt (2001), Waldman (2003), Krakel and Schottner (2012), and Auriol, Friebel, and von Bieberstein (2013) provide incentive-based explanations; and Prendergast (1993), Demougin and Siow (1994), Gibbons and Waldman (1999a), Zabochnik and Bernhardt (2001), Camara and Bernhardt (2009), and DeVaro and Morita (2013) provide explanations based on human-capital acquisition.

effort (Lazear and Rosen, 1981; Malcomson, 1984; Rosen, 1986; MacLeod and Malcomson, 1988; Prendergast, 1993; Chan, 1996; Zabojnik and Bernhardt, 2001; Waldman, 2003), it would be straightforward to extend them to generate the result that there is insider bias in hiring at the top. Only two of these papers (Chan, 1996; Waldman, 2003) specifically discuss insider bias in hiring at the top, but the main mechanisms of all these papers are consistent with this practice.

Existing papers, however, have paid less attention to the idea that promotion opportunities are constrained by a firm’s organizational structure (with the exception of Waldman (2003) and DeVaro and Morita (2013)), and none have considered how firms may put in place personnel policies to relax these constraints. As a result, one of our main contributions is to draw out the implication that firms may adopt forced-turnover policies specifically to create promotion opportunities. Taken together, there are many models of internal labor markets, and some of them are consistent with a subset of the features we have highlighted, but none is consistent with all of them.

Firm-Level Characteristics and Personnel Policies. Our approach not only shows that specific personnel policies might arise, but it also shows when they might arise. That is, our model describes how firm-level characteristics impact the types of personnel policies firms put in place.

In particular, our model suggests that firms with different spans—the ratio of the number of bottom jobs to top jobs—will put in place different personnel policies. We will focus on two such policies: insider bias in hiring at the top and forced-turnover policies. In addition, we briefly discuss the relationship between a firm’s span and its wage and promotion policies.

First, firms with larger spans will exhibit more of an insider bias in hiring at the top. To see this, suppose the firm has a one-time opportunity to hire into the top job an external candidate who will increase the NPV of future revenues by Δ . If the firm hires this external worker, the firm has to offer him total rents of v_2^* . If the firm instead hires into the bottom job and promotes a bottom worker, the firm will pay total rents of $v_1^* = 0$. As a result, the firm will hire the external candidate into the top job only if $\Delta > v_2^*$. By Proposition 1, under optimal personnel policies, firms with larger spans offer top workers greater rents, v_2^* .

Second, firms with larger spans will be more likely to put in place forced-turnover policies, since such firms will naturally have more limited promotion opportunities. In contrast, alternative justifications for mandatory-retirement policies, such as Lazear’s (1979) description of mandatory retirement as a provision in an optimal long-term contract, point to worker-level characteristics, rather than firm-level organizational characteristics, as the key drivers of such policies.

There is very little direct evidence that we are aware of documenting the relationship between

a firm’s span and its personnel policies, in large part because data on firms’ spans are limited.⁶ Nevertheless, we can take advantage of the empirical pattern that larger firms tend to have larger spans (Colombo and Delmastro (1999) and Caliendo, Monte, and Rossi-Hansberg (Forthcoming)) to provide indirect evidence. Recent evidence suggests that larger firms are indeed more likely to bias hiring into high-level positions towards insiders (DeVaro and Morita (2013) and Bond (2015)). And, according to a U.S. Department of Labor report, larger firms are more likely to put in place mandatory-retirement policies: “When firms were asked their reasons for using mandatory retirement, all firms, but particularly large firms, put greatest emphasis on assuring promotion opportunities for younger workers.” (U.S. Department of Labor, 1981).⁷

Third, larger firms will tend to pay all workers more at a given level in the firm, and workers in large firms will have lower promotion rates, as we show in Corollary 5. Barron, Black, and Loewenstein (1987) and Brown and Medoff (1989) find that larger firms offer higher starting wages. Our prediction about promotion rates is consistent with several isolated studies in the law industry, but we are unaware of any cross-industry evidence. Rebitzer and Taylor (1995) find that larger law firms offer higher wages both to their associates and to their partners, and more recently Garicano and Hubbard (2009) find that larger law firms offer higher pay to their top lawyers. Galanter and Palay’s (1991) broad study of law firms notes that “the chances of promotion to partner are accordingly lower in big firms than small firms.”

Our approach therefore highlights how firms’ characteristics affect workers’ careers through their impact on personnel policies. In particular, our results imply that workers’ careers in larger firms will be more likely to start at the bottom, where they will be relatively better paid, but their promotion prospects will be more limited, and their jobs will be less secure after promotion.

7 Extensions

In this section, we explore several extensions of the main model. The first extension shows that the optimal personnel policies described in Proposition 1 can emerge as part of an industry equilibrium when there are multiple competing firms. In the second extension, we study optimal personnel policies when wage policies and assignment policies can be history-dependent. The third extension discusses how the model can be extended to account for a richer set of stylized facts concerning wage dynamics within internal labor markets.

⁶Two exceptions include a pair of secondary findings by Guadalupe and Wulf (2010) and DeVaro and Kauhanen (2015). The former provide evidence that firms with larger spans pay higher wages to managers, controlling for firm size, and the latter find that promotion rates are lower in firms with larger spans.

⁷We thank Michael Waldman for this quote.

7.1 Industry Equilibrium

In the main model, a single firm operates in isolation, and the outside options of workers are exogenously specified. In this extension, we consider an environment with multiple competing firms in order to endogenize workers' outside options, and we show that the results of our main model remain unchanged. Optimal personnel policies possess the features described in Proposition 1, even when workers and firms earn no rents. For expositional clarity, we make a number of strong assumptions, but we discuss how the main results are preserved under weaker assumptions.

Suppose there is a mass I of identical firms and a mass L of identical workers. At the beginning of each period t , each worker is either employed by a firm or unemployed. If he is unemployed, he may receive a job offer from one of the firms, and he can accept the offer or reject it. If he rejects the offer or does not receive an offer, he gets a payoff of 0 for the period and remains unemployed at the beginning of the next period. If he accepts the offer, he works for the firm, and production takes place. We focus on symmetric, stationary industry equilibria in which firms adopt identical personnel policies and organizational structures, and these choices maximize their steady-state profits.

Each offer is characterized by (w_i, p_{ij}) , $i, j = 1, 2$. Let V_U denote the total utility of an unemployed worker, and let V_E^i , $i = 1, 2$, denote the total value of a worker working in activity i today. Let U be the mass of unemployed workers and D_i , $i = 1, 2$, be the mass of vacancies in activity i at the beginning of each period. We assume that each unemployed worker receives a job offer in activity i with probability $\alpha_i = D_i/U$. Finally, each firm's steady-state profits are given by $F(N_1, N_2) - w_1N_1 - w_2N_2$.

PROPOSITION 2. *There exists L^* such that if $L > L^*$, then $D_2 = 0$. In addition, $V_U = V_E^1 = 0$. Each firm's optimal personnel policies are the same as those described in Proposition 1, and the results of Corollaries 2-5 continue to hold.*

Proposition 2 shows that if workers are on the long side of the market, then if each firm believes that each other firm is adopting the personnel policies described in Proposition 1, it is optimal to do so as well. A key feature of Proposition 2 is that the total utility of an unemployed worker is the same as the total utility of a worker who is permanently unemployed. To see this, recall that in the main model, new workers earn zero rents under the optimal personnel policy when there is a single firm. With multiple firms, each firm will also find it optimal to make the same offer as other firms, leaving new workers with zero rents.

Since workers earn zero rents, unemployment does not serve as a worker discipline device for

bottom workers, as it would in Shapiro and Stiglitz’s (1984) model. Bottom workers are motivated to exert effort not because they fear losing their job, but rather because they value the prospect of eventually being promoted. For workers in the top job, unemployment still serves as a discipline device. Nevertheless, the unemployment rate no longer matters, because once workers lose their job, they are indifferent about how long it will take to find a new job.

The zero-rent feature also implies that workers’ outside options are essentially exogenous. As a result, the main results are preserved under more general settings. For example, the per-period payoff of an unemployed worker can be any value b . In this case, optimal personnel policies simply adjust by increasing the wages (for both activities) by b . The main results would remain unchanged if there is firm-level heterogeneity, and firms can choose different personnel policies and organizational structures. In this case, workers’ outside options remain exogenous to any individual firm, since they are unaffected by the hiring decisions of any individual firm. Put differently, bottom workers’ outside options serve as a sufficient statistic for the competitive environment that each individual firm faces.

Finally, if firms produce identical final goods for which there is a downward-sloping aggregate demand curve, then we can introduce a fixed entry cost and allow for endogenous firm entry subject to a zero ex-ante profit condition. In this case, firms would put in place the personnel policies described in Proposition 1, and all workers and firms would earn zero ex-ante rents.

7.2 Pay-for-Performance Contracts

In our main model, we made the stark assumption that wages are tied to activities. This assumption rules out both performance pay and other more flexible arrangements such as seniority-based raises and promotions. In this extension, we expand the firm’s contracting possibilities by allowing firms to put in place history-contingent pay-for-performance contracts and assignment policies. To maintain the idea that backloading pay within an activity 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 wage $w_t \geq \underline{w} \geq 0$ at the beginning of each period and a performance-contingent bonus $b_t \geq 0$ after each workers’ performance signal has been realized,

but before voluntary departures occur.⁸ 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 $h^t = (h_1, \dots, h_t)$, where $h_s \in \{1, 2\}$, $s = 1, 2, \dots, t$, denotes the activity to which the worker was assigned in period s . We therefore denote the worker's wage and bonus in period t by $w(h^t)$ and $b(h^t)$. We also allow the firm's assignment policy to depend on h^t . Denote by $p_i(h^t)$, $i = 1, 2$, 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 $\{w(h^t), b(h^t), p_1(h^t), p_2(h^t)\}_{t=1}^{\infty}$.

Since compensation and activity assignment 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 compensation policy, the firm is not restricted to setting the same wage for all workers performing the same activity. Wages can, for example, increase with the time on the activity. A worker's bonus can depend both on the time he has spent on the activity and on the time he has been with 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 compensation 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

⁸The analysis is essentially unchanged if bonus payments are delayed to the beginning of next period and become a part of the wage. That is, bonuses are not necessary, as long as wages are history-dependent. Allowing for bonus payments, however, simplifies our formal statements about the optimal personnel policies, and it allows us to highlight how the structure of pay-for-performance contracts depends on firm-level characteristics.

bound, so this contract is optimal. Proposition 2 describes the optimal contract.

PROPOSITION 3. *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 $d_2 N_2 / ((1 - d_1) N_1)$. Top workers perform Activity 2 and are never demoted. Workers are not fired unless they are caught shirking. (iii) The wage can be set to \underline{w} for all jobs. 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; 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 3 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 total 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 size 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.

Proposition 3 shows that the main features of the optimal personnel policies are mostly preserved. Workers enter the firm from the bottom job, and promotions are used to motivate them. In addition, the opportunities for promotion are key to the design of personnel policies. Proposition 3 also shows that the degree of flexibility in pay is important for determining optimal personnel policies when promotion opportunities are limited. In particular, firms need not put in place forced-turnover policies as long as bonus payments are feasible.

A feature of this extension is that even if pay-for-performance contracts are feasible, firms prefer to use promotions to motivate the bottom workers, and they will resort to bonuses only when promotion prospects are limited. Promotions arise as a cost-effective instrument for motivating workers in our model, because they allow the firm to optimally allocate separation rents from turnover at the top of the firm. Only when these separation rents are exhausted will firms complement promotions with additional bonuses at the bottom.

7.3 Internal Labor Markets

In Section 6, we focused on empirical regularities of internal labor markets that are consistent with our main model. In this subsection, we identify and discuss three additional empirical regularities that are not captured by our main model, and we outline how our model could be enriched to account for these patterns.

First, in our model, bottom workers are never fired, yet a significant amount of involuntary departures typically does occur in lower-level jobs within firms. Bottom workers are not fired in our model, in part because all workers are homogeneous, yet in practice, many workers may be a bad match for the firm at which they are employed (see Jovanovic (1979) for seminal work on match quality and worker turnover). One way to enrich our model to account for involuntary turnover at the bottom is to allow for heterogeneity in worker–firm match quality. For example, suppose each worker can either be a good or a bad match for the firm, but that this match quality is not known before he is hired, and once the worker joins the firm, his match quality is revealed. In such a model, bottom workers will be fired if they are revealed to be a bad match.

Additional regularities that our model does not explain are related to how wages change over time within jobs, how these wage changes depend on worker performance, and how these changes

are related to the likelihood of future promotions. Baker, Gibbs, and Holmstrom (1994a) report several key facts concerning performance, wage changes, and promotions. First, better performance is typically associated with a higher probability of promotion. Second, wages within jobs tend to increase with tenure on the job. Third, larger wage increases at one level of the job ladder are associated with a higher probability of quick promotion to the next level.⁹

To account for these facts, we can enrich our model by introducing symmetric learning about worker ability. Suppose there are two types of workers, high and low, and production efficiency requires that only the high-type worker can be productively assigned to activity 2. Worker types are not known *ex ante*, and they are learned only through workers' performance. If a worker shirks, bad performance occurs with some probability. When this occurs, the firm fires the worker. If a worker exerts effort, and if he is a low type, then performance is always at some intermediate level. If a high-type worker exerts effort, then performance is high with some probability and intermediate with the complementary probability. Worker performance is publicly observed, and high-type workers are valuable at other firms as well.

In this extension, a worker is revealed to be a high type when high performance is observed. In this case, the worker will be promoted to the top job if there are vacancies at the top. Otherwise, the worker will be put in a queue (with other workers who have been revealed to be high types) and will be promoted in the future. It then follows immediately that better performance is typically associated with a higher probability of future promotion: workers will not be promoted unless their performance is high.

In addition, wages also increase with tenure in the bottom job in this enrichment, and the reason for wage increases differs depending on whether workers have had high performance in the past. For workers who have not yet had high performance, their probability of being a high type falls over time, so promotions become less likely. To continue to motivate such workers, future wages must increase. For workers who have delivered high performance, their types are revealed to be high, increasing their outside option. The firm must then increase their wages in the bottom job in order to retain them, while they await promotion. To the extent that such wage increases are larger than the wage increases experienced by low-type workers, a larger wage increase is positively related to a higher probability of promotion to the next level.

⁹These findings have been confirmed by several follow-up studies (Seltzer and Merrett, 2000; Treble et al., 2001; Kauhanen and Napari, 2012).

8 Conclusion and Discussion

This paper develops a framework for studying how firms manage workers' careers in the presence of contractual imperfections. Under optimal personnel policies, low-incentive-rent activities are performed in entry-level jobs, and workers are motivated in part by the opportunity to advance to jobs requiring, and therefore delivering, greater 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. Firms may also optimally alter their organizational structures, becoming more top-heavy, in order to expand promotion opportunities.

Our framework generates novel testable implications linking firm characteristics to the personnel policies they put in place, and it has implications for how workers' careers might differ depending on the firm they work for. Firms with larger spans pay higher wages at all job levels, they have greater insider bias in hiring into top positions, and they may put in place forced-turnover policies such as mandatory-retirement policies. This implies that workers' careers in such firms will be more likely to start at the bottom, where they will be relatively better paid, but their promotion prospects will be more limited, and their jobs will be less secure after promotion.

Given its tractability, our model can serve as a building block for studying many further applications. For example, our model highlights how firms' personnel policies may be shaped by the policy environment in which the firm operates. As such, our model can be enriched to help us better understand the impacts of labor-market policies on workers' careers through their effects on firms' personnel policies. Preliminary analysis suggests that progressive taxation, which disproportionately affects top workers, also has indirect effects on bottom workers—fewer workers are hired at the bottom, but the workers who are hired have greater promotion opportunities. An increase in the minimum wage 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.

The economic forces we highlight in our model can be applied to study the design of more complicated hierarchies. As an example, Baron and Bielby (1986) show that job titles often do not correspond to different uses of the underlying production and are sometimes created as a way of sidestepping formal and rigid wage schedules.¹⁰ In our model, creating different job titles, with different pay levels, for the same activity allows the firm to backload pay within an activity as

¹⁰We thank Hideshi Itoh for suggesting this reference.

well as across activities. This allows the firm to extract rents from the worker more efficiently. Moreover, the benefit of backloading pay within jobs is greater when backloading pay across jobs is more costly. Our model therefore suggests that the benefits of job-title proliferation are greater 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.

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 its 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-For Online Publication

LEMMA 1. *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 $F_i(N_i^{pc}, N_j^{pc}) = w_i^{pc} > c_i$.*

PROOF OF LEMMA 1. 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

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

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 2-4, 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 2. *Assume $v_1^* \leq v_2^*$. Then $H_2^* = 0$.*

PROOF OF LEMMA 2. 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 3. Assume $v_1^* \leq v_2^*$ and $(1 - d_1)N_1R_1 \leq d_2N_2R_2$. Then $v_1^* = 0$, $v_2^* = R_2$, $p_{21}^* = 0$ and $p_{i1}^* + p_{i2}^* = 1$ for $i = 1, 2$.

PROOF OF LEMMA 3. 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 $d_2N_2R_2 \geq (1 - d_1)N_1R_1$. 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 2 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 4. Assume $v_1^* \leq v_2^*$ and $(1 - d_1)N_1R_1 > d_2N_2R_2$. Then $v_1^* = 0$, $v_2^* > R_2$, $p_{21}^* = 0$, $p_{22}^* < 1$, and $p_{11}^* + p_{12}^* = 1$.

PROOF OF LEMMA 4. Define $\Delta_i \equiv p_{i1}v_1 + p_{i2}v_2 - R_i$ to be the excess rents provided to workers performing activity i . We can write

$$\begin{aligned} M_1\Delta_1 + M_2\Delta_2 &= (M_1p_{11} + M_2p_{21})v_1 + (M_{12}p_{12} + M_2p_{22})v_2 - M_1R_1 - M_2R_2 \\ &= (N_1 - H_1)v_1 + (N_2 - H_2)v_2 - M_1R_1 - M_2R_2, \end{aligned}$$

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

$$\begin{aligned} W &= N_1c_1 + N_2c_2 + ((1 - \delta)N_1 + \delta H_1)v_1 + ((1 - \delta)N_2 + \delta H_2)v_2 \\ &= N_1c_1 + N_2c_2 + (1 - \delta)(M_1R_1 + M_2R_2) + H_1v_1 + H_2v_2 + (1 - \delta)(M_1\Delta_1 + M_2\Delta_2). \end{aligned}$$

Note that this implies that

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

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 3 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 2-4 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 4 that one lower bound for the wage bill (that holds whether $v_1^* \leq v_2^*$ or $v_1^* > v_2^*$) is:

$$W \geq N_1c_1 + N_2c_2 + (1 - \delta)(M_1R_1 + M_2R_2).$$

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

$$\begin{aligned} W &= N_1c_1 + N_2c_2 + v_1((1 - \delta)N_1 + \delta H_1) + v_2((1 - \delta)N_2 + \delta H_2) \\ &\geq N_1c_1 + N_2c_2 + (1 - \delta)N_1R_2, \end{aligned}$$

where the expression for the wage bill (that holds whether $v_1^* \leq v_2^*$ or $v_1^* > v_2^*$) is obtained in Lemma 2. This is our second lower bound.

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

$$W \geq N_1c_1 + N_2c_2 + (1 - \delta) \max \{M_1R_1 + M_2R_2, N_1R_2\}$$

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 $N_1R_2 \leq M_1R_1 + M_2R_2$. In this case, let $v_1^* = (M_1R_1 + M_2R_2)/N_1$ and $v_2^* = 0$. Let $p_{11}^* = N_1R_1/(M_1R_1 + M_2R_2)$, $p_{12}^* = 0$, $p_{21}^* = N_1R_2/(M_1R_1 + M_2R_2)$, and $p_{22}^* = 1 - p_{21}^*$. Notice that these probabilities are less than 1 because $M_1R_1 + M_2R_2 \leq N_1R_2$. In addition, $(IC - 1)$, $(IC - 2)$, $(IR - 1)$, and $(IR - 2)$ are satisfied given the set of choices. Now let $H_1^* = 0$ and $H_2^* = N_2d_2 + N_2(1 - d_2)p_{21}^*$. The flow constraints are satisfied. Finally, the wage bill associated with this policy is given by $N_1c_1 + N_2c_2 + (1 - \delta)(M_1R_1 + M_2R_2)$, which reaches the lower bound.

Next, consider the case with $N_1R_2 > M_1R_1 + M_2R_2$. 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 $N_1R_2 > M_1R_1 + M_2R_2$, we have $N_1R_2 > M_2R_2$, 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 $N_1c_1 + N_2c_2 + (1 - \delta)N_1R_2$, 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

$$W|_{v_1^* > v_2^*} = \begin{cases} N_1c_1 + N_2c_2 + (1 - \delta)(M_1R_1 + M_2R_2) & \text{if } N_1R_2 \leq M_1R_1 + M_2R_2 \\ N_1c_1 + N_2c_2 + (1 - \delta)N_1R_2 & \text{if } N_1R_2 > M_1R_1 + M_2R_2, \end{cases}$$

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

Next, to characterize the optimal personnel policy, recall from Lemmas 2-4 that when $v_1^* \leq v_2^*$, the total wage bill is given by

$$W|_{v_1^* \leq v_2^*} = \begin{cases} N_1c_1 + N_2c_2 + (1 - \delta)(M_1R_1 + M_2R_2) & \text{if } N_2R_2 \leq M_1R_1 + M_2R_2 \\ N_1c_1 + N_2c_2 + (1 - \delta)N_2R_2 & \text{if } N_2R_2 > M_1R_1 + M_2R_2. \end{cases}$$

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

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

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

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

(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 $MRTS(N_1, N_2) > W_1^-(N_1, N_2)/W_2^-(N_1, N_2)$, 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 :

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

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

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

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

COROLLARY 4. *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 4. 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 $(N_1^*(y^*), \xi_{W(y^*)}(N_1^*(y^*)))$ 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 $\lim_{y^* \rightarrow 0} MRTS(N_1^*(y^*), \xi_{W(y^*)}(N_1^*(y^*))) = (c_1 / (c_2 + (1 - \delta)R_2))$. Along the ray $(\kappa N_2, N_2)$, there exists some \hat{N}_2 such that for all $N_2 < \hat{N}_2$,

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

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

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

Define y_1 to be the supremum over all y such that

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

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

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

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

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

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

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

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

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

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

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

COROLLARY 5. *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 5. 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 $w_2^{*S} = c_2/((1 - d_2)\delta)$. Further, the promotion rate for bottom workers is $p_{12}^{*S} = d_2 N_2/((1 - d_1)N_1)$, 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 $w_2^{*L} = (c_1 N_1 + c_2 N_2)/(\delta N_2)$. Further, by Lemma 4, 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 exists L^* such that for all $L > L^*$, we have $D_2 = 0$. In addition, $V_U = V_E^1 = 0$. The optimal personnel policy is the same as is described in Proposition 1, and the results of Corollaries 2-5 continue to hold.*

PROOF OF PROPOSITION 2. Recall that V_U is the NPV of an unemployed worker and V_E^i , $i = 1, 2$, are the NPV of a worker in activity i . It follows that V_U satisfies the following:

$$V_U = \alpha_1 V_E^1 + \alpha_2 V_E^2 + (1 - \alpha_1 - \alpha_2) \delta V_U.$$

Since each firm is infinitesimal, they take V_U as given when they maximize their profits. In other words, the firm maximizes

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

subject to

$$\begin{aligned} V_E^1 &= w_1 - c_1 + (1 - d_1) \delta (p_{11}V_E^1 + p_{12}V_E^2) + \delta (d_1 + (1 - d_1) (1 - p_{11} - p_{12})) V^U \quad (\text{PK-1-m}) \\ V_E^2 &= w_2 - c_2 + (1 - d_2) \delta (p_{21}V_E^1 + p_{22}V_E^2) + \delta (d_2 + (1 - d_2) (1 - p_{11} - p_{12})) V^U \quad (\text{PK-2-m}) \end{aligned}$$

$$V_E^1 \geq V^U; \quad (\text{IR-1-m})$$

$$V_E^2 \geq V^U. \quad (\text{IR-2-m})$$

$$p_{11} (V_E^1 - V^U) + p_{12} (V_E^2 - V^U) \geq c_1 / ((1 - d_1) \delta q_1) = R_1; \quad (\text{IC-1-m})$$

$$p_{21} (V_E^1 - V^U) + p_{22} (V_E^2 - V^U) \geq c_2 / ((1 - d_2) \delta q_2) = R_2, \quad (\text{IC-2-m})$$

$$(1 - d_1) p_{11} N_1 + (1 - d_2) p_{21} N_2 + H_1 = N_1; \quad (\text{FL-1-m})$$

$$(1 - d_1) p_{12} N_1 + (1 - d_2) p_{22} N_2 + H_2 = N_2, \quad (\text{FL-2-m})$$

where $H_i \geq 0$ is the mass of new workers hired into activity i . In addition, we have

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

Notice that by redefining $\hat{v}_i = V_E^i - V^U$, $i = 1, 2$, the constrained maximization problem above becomes the same as the constrained maximization problem in the main model, so it has the same solution as that in Proposition 1. In particular, we have $H_2^* = 0$, and $\hat{v}_1^* = 0$. This implies that $D_2 = IH_2^* = 0$, and thus, $\alpha_2 = 0$, and in addition, $V_E^1 = V^U$. We then have that

$$\begin{aligned} V_U &= \alpha_1 V_E^1 + \alpha_2 V_E^2 + (1 - \alpha_1 - \alpha_2) \delta V_U \\ &= \alpha_1 V_U + (1 - \alpha_1) \delta V_U. \end{aligned}$$

This gives that $V_U = V_E^1 = 0$.

Finally, we need to check that $\alpha_1 = U/D_1 < 1$. Now let N_1^* and N_2^* be the mass of positions, H_1^* be the mass of new hires, and $d_2^* = d_2 + (1 - d_2) (1 - p_2^{1*} - p_2^{2*})$ be the total turnover rate from the top job in each firm. This implies that the total vacancy is $D_1 = IH_1^* = I (d_1 N_1^* + d_2^* N_2^*)$. The total unemployment is given by $U = L - I (N_1^* + N_2^*) + I (d_1 N_1^* + d_2^* N_2^*)$. Therefore, as long as $L > I (N_1^* + N_2^*) \equiv L^*$, we have $U > D_1 = D$, so that $\alpha_1 < 1$. This proves the proposition. ■

PROPOSITION 3. *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 $d_2 N_2 / ((1 - d_1) N_1)$. 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.*

PROOF OF PROPOSITION 3. 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

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

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

and

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

Note that for any compensation plan with wage $\tilde{w}(h^t)$ and bonus $\tilde{b}(h^t)$, the firm can choose an alternative compensation plan with wage \underline{w} and bonus $\tilde{w}(h^t) + \tilde{b}(h^t) - \underline{w}$. 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 $w(h^t) = \underline{w}$ following all histories.

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

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

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

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

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

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

This gives the incentive constraint $(IC - h^t)$. 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 $L(h^t)$ as the mass of workers with history h^t . The flow constraint following history h^t can then be written as

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

In addition, there are two aggregate flow constraints:

$$\sum_{h^t|h_t=1} L(h^t) = N_1; \quad (\text{FL-1})$$

$$\sum_{h^t|h_t=2} L(h^t) = N_2. \quad (\text{FL-2})$$

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

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

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

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

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

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

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

This implies that

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

where the second equality follows because

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

and the last equality follows because

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

To establish a lower bound for $\sum_{h^t} L(h^t) b(h^t)$, below we separately provide a lower bound for $\sum_{h^t} L(h^t) v(h^t)$ 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

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

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

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

Thus, if we can choose a set of feasible $\{b(h^t), p_1(h^t), p_2(h^t)\}_{t=1}^\infty$ such that $\sum_{h^t} L(h^t) b(h^t)$ 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

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

where b_2 is given by

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

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

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

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

where

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

and

$$b_1 = c_1 + r_1 - \delta(1 - d_1)(pr_2 + (1 - p)r_1 + \underline{w}).$$

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

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

and

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

Therefore, it follows

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

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

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

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

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 $v(h^t, 1) \leq \underline{w} + r_2$ and $v(h^t, 2) = \underline{w} + r_2$ 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

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

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

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

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

Now suppose that $v(h^t, 2) > \underline{w} + r_2$ so that the incentive constraint is slack. It follows that if $b(h^t, 2) > 0$, the firm can use the type of perturbation above on $b(h^t, 2)$ and $b(h^t)$ since the incentive constraint on $(h^t, 2)$ is slack. If $b(h^t, 2) = 0$, the promise-keeping condition on $v(h^t, 2)$ (since it exceeds $\underline{w} + r_2$) implies that there either exists some $b(h^\tau|h^t, 2) > 0$ with $h_\tau = 1$ or some $b(h^{\tau'}|h^t, 2) > 0$ with $h_{\tau'} = 2$ and $v(h^{\tau'}|h^t, 2) > \underline{w} + r_2$. Otherwise, it contradicts the assumption that $v(h^t, 2) > \underline{w} + r_2$. In the former case, the firm can perform a feasible perturbation on $b(h^\tau|h^t, 2)$ and $b(h^t)$ 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 $b(h^{\tau'}|h^t, 2)$ and $b(h^t)$. This reduces the firm's total bonus payout, and therefore, contradicts the optimality of the original contract.

Similarly, consider the case with $v(h^t, 1) > \underline{w} + r_2$. If $b(h^t, 1) > 0$, then the firm can use the perturbation above on $b(h^t, 1)$ and $b(h^t)$. If $b(h^t, 1) = 0$, Assumption 3 then implies that there either exists some $b(h^\tau|h^t, 1) > 0$ with $h_\tau = 1$ or some $b(h^{\tau'}|h^t, 1) > 0$ with $h_{\tau'} = 2$ and $v(h^{\tau'}|h^t, 1) > \underline{w} + r_2$. By performing the perturbation on $b(h^t)$ and $b(h^\tau|h^t, 1)$ (or on $b(h^t)$ and $b(h^{\tau'}|h^t, 1)$), 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 $v(h^t, 1) \leq \underline{w} + r_2$ and $v(h^t, 2) = \underline{w} + r_2$.

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

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

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

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 GIBBS, AND BENGT HOLMSTROM. 1994a. The Internal Economics of the Firm: Evidence from Personnel Data. *Quarterly Journal of Economics*, 109(4): 881-919.
- [5] BAKER, GEORGE, MICHAEL GIBBS, AND BENGT HOLMSTROM. 1994b. The Wage Policy of a Firm. *Quarterly Journal of Economics*, 109(4): 921-955.
- [6] BARON, JAMES AND WILLIAM BIELBY. 1986. The Proliferation of Job Titles in Organizations. *Administrative Science Quarterly*, 31(4): 561-586.
- [7] 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.
- [8] BERNHARDT, DAN. 1995. Strategic Promotion and Compensation. *Review of Economic Studies*, 62(2): 315-339.
- [9] BOARD, SIMON. 2011. Relational Contracts and the Value of Loyalty. *American Economic Review*, 101(7): 3349-3367.
- [10] BOLTON, PATRICK AND MATHIAS DEWATRIPONT. 2005. *Contract Theory*. MIT Press.
- [11] BOND, TIMOTHY. 2015. Internal Labor Markets in Equilibrium. *Working Paper*.
- [12] BOSE, GAUTAM AND KEVIN LANG. 2013. A Theory of Monitoring and Internal Labor Markets. *Working Paper*.
- [13] BROWN, CHARLES AND JAMES MEDOFF. 1989. The Employer Size Wage Effect. *Journal of Political Economy*, 97(5): 1027-1059.
- [14] CALIENDO, LORENZO, FERDINANDO MONTE, AND ESTEBAN ROSSI-HANSBERG. Forthcoming. The Anatomy of French Production Hierarchies. *Journal of Political Economy*.
- [15] 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.
- [16] CAMARA, ODILON AND DAN BERNHARDT. 2009. The Dynamics of Promotions, Quits, and Layoffs. *Working Paper*.
- [17] CAPPELLI, PETER. 2008. *Talent on Demand*. Harvard Business Press.

- [18] CARMICHAEL, LORNE. 1985. Can Unemployment Be Involuntary?: Comment. *American Economic Review*, 75(5): 1213-1214.
- [19] CHAN, WILLIAM. 1996. External Recruitment versus Internal Promotion. *Journal of Labor Economics*, 14(4): 555-570.
- [20] COLOMBO, MASSIMO AND MARCO DELMASTRO. 1999. Some Stylized Facts on Organization and its Evolution. *Journal of Economic Behavior and Organization*, 40(3): 255-275.
- [21] DEMOUGIN, DOMINIQUE AND ALOYSIUS SIOW. 1994. Careers in Ongoing Hierarchies. *American Economic Review*, 84(5): 1261-1277.
- [22] DEVARO, JED AND ANTTI KAUKANEN. Forthcoming. An “Opposing Responses” Test of Classic Versus Market-Based Promotion Tournaments. *Journal of Labor Economics*.
- [23] DEVARO, JED AND HODAKA MORITA. 2013. Internal Promotion and External Recruitment: A Theoretical and Empirical Analysis. *Journal of Labor Economics*, 31(2): 227-269.
- [24] DOERINGER, PETER AND MICHAEL PIORE. 1971. *Internal Labor Markets and Manpower Analysis*. Heath Lexington Books.
- [25] DOHMEN, THOMAS. 2004. Performance, Seniority and Wages: Formal Salary Systems and Individual Earnings. *Labour Economics*, 11(6): 741-763.
- [26] FONG, YUK-FAI AND JIN LI. 2015. Relational Contracts, Limited Liability, and Employment Dynamics. *Working Paper*.
- [27] GALANTER, MARC AND THOMAS PALAY. 1991. *Tournament of Lawyers: The Transformation of the Big Law Firm*. University of Chicago Press.
- [28] GARICANO, LUIS AND THOMAS HUBBARD. 2009. Earnings Inequality and Coordination Costs: Evidence from U.S. Law Firms. *Working Paper*.
- [29] GHOSH, SUMAN AND MICHAEL WALDMAN. 2010. Standard Promotion Practices versus Up-or-out Contracts. *RAND Journal of Economics*, 41(2): 301-325.
- [30] GIBBONS, ROBERT. 1997. Incentives and Careers in Organizations. In *Advances in Econometrics: Theory and Applications*, eds. David Kreps and Kenneth Wallis, Cambridge University Press.
- [31] GIBBONS, ROBERT AND MICHAEL WALDMAN. 1999a. A Theory of Wage and Promotion Dynamics inside Firms. *Quarterly Journal of Economics*, 114(4): 1321-1358.
- [32] 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.
- [33] GRUND, CHRISTIAN. 2005. The Wage Policy of Firms: Comparative Evidence for the US and Germany from Personnel Data. *International Journal of Human Resource Management*, 16(1): 104-119.

- [34] GUADALUPE, MARIA AND JULIE WULF. 2010. The Flattening Firm and Product Market Competition: The Effect of Trade Liberalization on Corporate Hierarchies. *American Economic Journal: Applied Economics*, 2(4): 105-127.
- [35] HARRIS, MILTON AND BENGT HOLMSTROM. 1982. A Theory of Wage Dynamics. *Review of Economic Studies*, 49(3): 315-333.
- [36] HUSON, MARK, PAUL MALATESTA, AND ROBERT PARRINO. 2004. Managerial Succession and Firm Performance. *Journal of Financial Economics*, 74(2): 237-275.
- [37] JENSEN, MICHAEL. 1986a. Agency Costs of Free Cash Flow, Corporate Finance and Takeovers." *American Economic Review*, 76(2): 323-329.
- [38] JENSEN, MICHAEL. 1986b. The Takeover Controversy: Analysis and Evidence. *Midland Corporate Finance Journal*, 4(2): 6-32.
- [39] JOVANOVIC, BOYAN. 1979. Job Matching and the Theory of Turnover. *Journal of Political Economy*, 87(5): 972-990.
- [40] KAUKANEN, ANTTI AND SAMI NAPARI. 2012. Career and Wage Dynamics: Evidence from Linked Employer-Employee Data. *Journal of Labor Research*, 36: 35-76.
- [41] KRAKEL, MATTHIAS AND ANJA SCHOTTNER. 2012. Internal Labor Markets and Worker Rents. *Journal of Economic Behavior and Organization*, 84(2): 491-509.
- [42] KWON, ILLONG. 2006. Incentives, Wages and Promotions: Theory and Evidence. *RAND Journal of Economics*, 37(1): 100-120.
- [43] LAZEAR, EDWARD. 1979. Why is there Mandatory Retirement? *Journal of Political Economy*, 87(6): 1261-1284.
- [44] 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.
- [45] LAZEAR, EDWARD. 1999. *Personnel Economics: Past Lessons and Future Directions*. NBER Working Paper w6957.
- [46] LAZEAR, EDWARD AND PAUL OYER. 2013. *Personnel Economics*. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.
- [47] LAZEAR, EDWARD AND SHERWIN ROSEN. 1981. Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, 89(5): 841-864.
- [48] LAZEAR, EDWARD, KATHRYN SHAW, AND CHRISTOPHER STANTON. Forthcoming. Making Do With Less: Working Harder During Recessions. *Journal of Labor Economics*.
- [49] LIN, MING-JEN. 2005. Opening the Black Box: The Internal Labor Markets of Company X. *Industrial Relations*, 44(4): 659-706.

- [50] MACLEOD, BENTLEY AND JAMES MALCOMSON. 1988. Reputation and Hierarchy in Dynamic Models of Employment. *Journal of Political Economy*, 96(4): 832-854.
- [51] MALCOMSON, JAMES. 1984. Work Incentives, Hierarchy, and Internal Labor Markets. *Journal of Political Economy*, 92(3): 486-507.
- [52] MCCUE, KRISTIN. 1996. Promotions and Wage Growth. *Journal of Labor Economics*, 14(2): 175-209.
- [53] MILGROM, PAUL. 1988. Employment Contracts, Influence Activities, and Efficient Organizational Design. *Journal of Political Economy*, 96(1): 42-60.
- [54] MILGROM, PAUL AND JOHN ROBERTS. 1988. An Economic Approach to Influence Activities in Organizations. *American Journal of Sociology*, 94: S154-S179.
- [55] MOOKHERJEE, DILIP. 2013. Incentives in Hierarchies. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.
- [56] MURPHY, KEVIN. 1985. Corporate Performance and Managerial Remuneration: An Empirical Analysis. *Journal of Accounting and Economics*, 7(1): 11-42.
- [57] OYER, PAUL. 2007. Is There an Insider Advantage in Getting Tenure? *American Economic Review*, 97(2): 501-505.
- [58] PRENDERGAST, CANICE. 1993. The Role of Promotion in Inducing Specific Human Capital Acquisition. *Quarterly Journal of Economics*, 108(2): 523-534.
- [59] QIAN, YINGYI. 1994. Incentives and Loss of Control in an Optimal Hierarchy. *Review of Economic Studies*, 61(3): 527-544.
- [60] 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.
- [61] ROSEN, SHERWIN. 1986. Prizes and Incentives in Elimination Tournaments. *American Economic Review*, 76(4): 701-715.
- [62] SELTZER, ANDREW 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.
- [63] SHAPIRO, CARL AND JOSEPH STIGLITZ. 1984. Equilibrium Unemployment as a Worker Discipline Device. *American Economic Review*, 74(3): 433-444
- [64] TREBLE, JOHN, EDWIN VAN GAMEREN, SARAH BRIDGES, AND TIM BARMBY. 2001. The Internal Economics of the Firm: Further Evidence from Personnel Data. *Labour Economics*, 8(5): 531-552.
- [65] U.S. DEPARTMENT OF LABOR. 1981. Abolishing Mandatory Retirement. Washington, U.S.: Government Printing Office.

- [66] WALDMAN, MICHAEL. 1984. Job Assignments, Signalling, and Efficiency. *RAND Journal of Economics*, 15(2): 255-267.
- [67] WALDMAN, MICHAEL. 2003. Ex Ante versus Ex Post Optimal Promotion Rules: The Case of Internal Promotion. *Economic Inquiry*, 41(1): 27-41.
- [68] WALDMAN, MICHAEL. 2013. Theory and Evidence in Internal Labor Markets. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.
- [69] WILLIAMSON, OLIVER. 1967. *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm*. Chicago, Illinois: Markham Publishing Company.
- [70] 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.