Managing Careers in Organizations

Preliminary - Please do not circulate without permission.

Rongzhu Ke Jin Li Michael Powell*

October 31, 2013

Abstract

We extend the classic Shapiro and Stiglitz model to allow for multiple jobs within an organization. Organizations make wage-, promotion-, recruiting-, and job-security-policy decisions to motivate workers. A hierarchical structure emerges: organizations institute a port of entry and a linear career progression. Workers at the bottom receive no rents. Instead, efficiency wages at the top motivate workers throughout the organization, and the organization may therefore promote turnover at the top to increase worker mobility through the ranks. These internal considerations distort production away from the technologically ideal. Finally, we examine the effects of policies including minimum wages, progressive taxation, retirement-program eligibility, and employment protection on careers.

*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 are grateful to for helpful suggestions.

1 Introduction

Career-based incentives such as promotions serve as an important motivator for workers. Actively managing vertical mobility–that is, actively managing careers within organizations is an important challenge for most firms. Failure to create enough promotion opportunity can demotivate the workers and hurt the bottom line. In fact, Cappelli (2008) suggests that "frustration with advancement opportunities is among the most important factors pushing individuals to leave for job elsewhere."

To manage careers of the workers, the firm has many levers to pull. The firm can design its recruiting policy by deciding how many workers to hire at the each level of the firm: hiring fewer external candidates boosts the promotion opportunity of internal candidates. The firm can also affect its promotion policy by adjusting the hierarchical structure of the organization: increasing the number positions at one level makes promotion into that level easier. In addition, the firm can design its retention policy: encouraging workers at some level of the hierarchy to exit means more positions become available. Moreover, the firm can design its wage policy by determining the pay-grade it attaches to each level of the organization: enlarging the pay difference between job levels strengthens incentives.

While all these policies have benefits in motivating the workers, they also impose costs in other dimensions. A recruiting policy that favors insiders limits the firm's ability to tap talent from the market place. A promotion policy that facilitates career advancement with a small hierarchical span may distort the production by having too many workers at the top. A harsh retention policy can force the firm to pay workers extra to compensate for their job insecurity. Finally, a wage policy that gives large wage increases upon promotion raises the total wage bill.

In general, all these policies are interdependent, and the effectiveness of one policy depends on how and what other policies are used. A key challenge for the firm is therefore to determine how to optimally combine the recruiting, promotion, retention, and the wage policies. These policies together form the core of a firm's human resources management policy and largely determine the careers of the workers within the organization. The purpose of this paper is to offer a simple and tractable framework to study how the firm chooses these policies jointly, and, thus, how the firm

can best manage the careers of its workers.

Our model shows that the optimal policies result in a number of salient employment patterns that resemble an internal labor market. In particular, there is a port of entry into the firm and workers within the firm gradually climb the job ladder until they exit. Beyond these qualitative patterns, the model allows us to study how the hierarchical structure, speed of career progression, exit rate, and wages change with respect to the production function, exogenous turnover rates, and outside options. A noticeable feature of the model is that when it is important to motivate workers at lower level jobs, the firm may hire more workers at the top job, paying them higher wage, and having a harsher retention rule than it would do otherwise.

Our baseline model builds on the celebrated efficiency-wage of Shapiro and Stiglitz (1984). As in Shapiro and Stiglitz, the workers privately choose whether to work or shirk. The firms cannot use contracts to motivate the workers but instead use the threat of firing those who are caught shirking. For the firing to be effective, the firms must pay the workers a wage above their outside options, and hence, the workers receive rents on the job. Different from Shapiro and Stiglitz, in our model each firm has two jobs. We denote the job with higher rent as the top job and the other one as the bottom job. The firm's output depends jointly on the number of top and bottom jobs, and the firms maximize their steady-state profits.

To do so, each firm decides on the number of top and bottom jobs to have and chooses the associated wages. Moreover, the firm chooses how many new workers to hire into each job in each period. For the existing workers, the firm decides whether to keep them, and if they are kept, which job to assign them to depending on their position in the last period. These decisions of the firm are limited by two key constraints. First, these decisions must induce workers to exert effort in each job. This is the standard incentive compatibility constraint for the worker. A less common constraint for this model is the flow constraint: for the firm to be in the steady state the number of workers that flow into each job must be equal to the number of workers flow out of each job.

A benchmark case for our setup is that firm can treat the two jobs separately, i.e., paying Shapiro-Stiglitz efficiency wage on each job. But this is suboptimal. In the optimal recruiting policy, the firm only hires workers into the bottom job (unless there are not enough existing workers

to fill the top job). In other words, there is a port of entry and the firm promotes within. By doing so, the firm lowers the rents given out: when a worker at the top job leaves, the rents associated with the job is not given to an outsider worker but an existing one, who "pays" for the promotion opportunity by accepting a lower wage for the bottom job. By promoting from within, the efficiency-wage rents associated with the top job are reused: these rents motivate the workers at the top and also serve as promotion incentives for workers at the bottom. As a result, the firm can increase its profit by reducing the rents at the bottom worker.

This advantage for promotion from within follow from the same logic of favoring insiders as a way of reusing the rents; see Board (2011) for a formal model in the context of supplier relationship. Relatedly, the idea that promotion can be used as a way to backload the payment so as to lower the rents to the worker is very related to the "Trust Funds" idea by Akerlof and Katz (1989). In our context, however, a surprising result is that the reusability of rents can be made so effective that the rents of the bottom job are always zero.

It is clear that when the rent of the top job is high enough, the payoff of the workers become sufficiently backloaded and there is no need to give rents to workers at the bottom job. However, when the rent of the top job is not too large or the exogenous turnover rate at the top job is not so high, the promotion incentive alone is not enough to motivate the workers at the bottom job. In this case, one might have thought that the firm can complement the promotion incentive with an efficiency-wage incentive for the bottom job, and, therefore, giving some rent to the workers.

This, however, is suboptimal. Instead, the firm should actively manage the exit rate of workers at the top to expand the promotion opportunity. In the optimal retention policy, the firm pushes out as few workers as needed so as to create just enough promotion opportunity for motivating the workers at the bottom job. Notice that pushing top workers is costly for the firm because the firm needs to pay top workers more to compensate for the lower job security. Nevertheless, this is still more profitable than paying efficiency wages to bottom workers: the firm recoups some of its cost of raise the wages of top workers by lowering the wages of bottom workers whereas it does not recoup its loss from paying higher wage to the bottom workers.

The zero-rent result for workers at the bottom implies that efficiency wages are not used for

bottom worker. Consequently, the wage gap between the two jobs within the internal labor market is strictly larger than if these two jobs are considered separately as in the benchmark Shapiro-Stiglitz case. In general, the availability of internal labor markets allows firms to partially back-load compensation, even when wages are contractually tied to jobs. More interestingly, when the Shapiro-Stiglitz rents of the top jobs are insufficient to motivate the bottom workers, the firm will both raise the wage and lower the job security. In this case, it can be shown that the wage increase upon promotion is increasing in the span of the hierarchy.

Notice that the size and span of the hierarchy is determined endogenously in our model, and they reflect of trade-off between productive efficiency and incentive provision. In particular, for a fixed number of workers at the top job, the number of bottom jobs can be efficient. This happens when exogenous turnover in the top job is sufficient to provide incentive at the bottom. When this is not the case, the number of bottom workers is inefficiently small even if each of them receives zero rents. The reason for the inefficiency is that there is spillover distortion at the top: for each additional hire at the bottom, the firm must lower the job security top, and, hence, increase the wage at the top. This inefficiency also implies that for a fixed number of top jobs, the number of workers at the bottom is weakly increasing in the turnover rate at the top. Essentially, a higher turnover rate at the top increases the promotion opportunity at the bottom, and this allows the firm to hire more workers at the bottom. Recall from the Shapiro and Stiglitz model that the turnover rate and the employment level are always negatively correlated. In this model, however, the turnover rate at the top can be positively correlated to the number of workers at the bottom.

Our baseline model is tractable enough that we are able to study the effects on internal labor markets from a number of labor market policies such as progressive tax, employment protection, and minimum wage. A key feature of our model is that the internal labor market is interconnected that policies with direct effects on a targeted group of the workers will lead to indirect effects on other groups of workers. In particular, we show that progressive tax policies that affect the wages of the top workers may lead to lower employment levels for the bottom workers although each employed bottom worker has a better chance of career advancement. In addition, we show that policies that ban firing of the workers will lead to both a lower level of total employment and also

a smaller span of the hierarchy. Finally, we show that minimum wage can sometimes increase the employment level of the firm.

The rest of the paper is organized as follows. Section 2 sets up the model and Section 3 describes the efficiency-wage benchmark. The main results of the model are described in Section 4 and Section 5 describes the effects of labor market policies. Section 6 concludes and discusses future extensions of the model.

Related Literature This paper contributes to a number of literatures. First, our paper is related to the literature on tournaments, in which promotions serve as the motivating example for the optimality of rank-order tournaments; see for example Lazear and Rosen (1981) for the seminal paper in this literature. We contribute to this literature by considering not only the motivational aspects of promotions as prizes but also the firm implements the promotion incentive by choosing both the wage upon promotion and the opportunity for promotion.

We also contribute to the efficiency-wage literature; see Akerlof and Yellen (1986) for a collection of important efficiency-wage models. One lesson that emerges from this literature is that backloaded pay schedules are typically beneficial to the firm. Our model formally analyzes the use of promotions as a device for backloading pay when wages are tied to jobs. We show that when pay is backloaded through promotions, all rents are extracted from incoming workers. Consequently, it would seem that, because workers are indifferent between taking a job and remaining unemployed, production would be undistorted in an economy. Nevertheless, we show that noncontractibility of effort still leads to productive inefficiency, since it distorts the relative number of positions at the top, relative to the bottom, of the organization. Further, our model sheds new light on the relationship between turnover and the employment level. In the classic Shapiro and Stiglitz model, the turnover rate is negatively associated with the employment level, because a higher turnover rate implies higher efficiency wages. In our model, a higher turnover rate at the top job can be positively associated with the number of workers at the bottom, because more turnover at the top increases the promotion opportunities for those at the bottom.

Finally, our paper is related to the literature on internal labor markets; see Gibbons (1997)

Gibbons and Waldman (1999), Lazear (1999), Lazear and Oyer (2013) and Waldman (2013) for reviews. We contribute to this literature by constructing an integrated model that helps explain a number of well-known facts on internal labor markets discussed in the seminal work by Doeringer and Piore (1971). Unlike most models in this literature, the voluntary turnover rate of workers plays a key role in our result. As Waldman (2013) notes in his recent review: "There are a few subjects that deserve more attention at both the theoretical and empirical levels. One subject is the connection between wage and promotion dynamics in internal labor markets and the turnover decision. How a worker's career progresses during a stay at a single employer should be closely related to voluntary and involuntary turnover decisions and also to how the worker performs at the new job." Our model fills this gap by serving as a framework to study how the promotion, job-security, hiring, and wage policies are jointly determined, and moreover, how these policies are influenced by exogenous environmental factors.

2 The Model

A single firm and a mass L of homogeneous workers interact repeatedly. Time is discrete and denoted by $t = 1, 2, ..,$ and the firm and workers share a common discount factor $\delta \in (0, 1)$. Throughout the analysis, we focus on the steady state and suppress time subscripts. Production in each period requires two tasks to be performed. Each worker is capable of performing either task if necessary, but he may perform only one task in a given period. A worker performing task 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**, and we refer to such a worker as **productive**. A worker's effort choice is his private information, but shirking in task i is contemporaneously detected with probability q_i. If in period t the firm employs masses N_1 and N_2 of productive workers in the two tasks, output is given by $f(N_1, N_2)$. We assume that L is large enough to accommodate any finite choice of N_1 and N_2 by the firm.

At the beginning of each period, the firm assigns existing workers to tasks. The firm then chooses a mass of new hires H_i and wage levels w_i for each task $i = 1, 2$. Throughout, as in Prendergast (1993), we assume that wages are tied to tasks. Workers may reject the wage in favor

of an outside option that yields utility 0. If they accept the offer, workers then choose effort levels, and wages are paid. At the end of the period, a fraction d_i of workers in task i exogenously leave the firm. We refer to d_i as the **exogenous turnover rate** of workers in task i. For those workers who do not exogenously leave the firm, the firm determines worker mobility patterns. To simplify notation, we assume without loss of generality that a worker who is found shirking is fired with probability 1. For the remaining workers, let p_{ij} denote the probability that a worker in task i will take on task j next period. We may have $p_{i1} + p_{i2} < 1$, and a worker not assigned to a task is released from the firm, receiving his outside utility.

At the end of period, workers who leave the firm also leave the workforce and are replaced by identical workers. We make this assumption at this point to maintain stationarity. In later sections, we relax this assumption and also allow workers' outside options to be determined endogenously in a market equilibrium.

3 Efficiency-Wage Benchmark

In order to provide a benchmark against which to compare our main results, we begin by describing what we will refer to as the **efficiency-wage benchmark**. In this benchmark, the firm treats the two tasks 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 cross-task mobility.

Given a mass $\hat{N}_{-i}$ of workers in task $-i$, the firm chooses N_i and w_i to solve the following program:

$$\max_{N_i, w_i} f\left(N_i, \hat{N}_{-i}\right) - w_i N_i$$

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

receives a total payoff of v_i in the job, where

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

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

The worker will choose high effort as long as

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

Choosing low effort allows the worker to avoid incurring the cost c_i but with probability q_i, the worker is caught shirking and fired. Wages, or equivalently, payoffs v_i, are reduced as much as possible to ensure that this constraint holds with equality. Given the resulting efficiency wage, the firm chooses the mass of workers such that the marginal benefit of hiring an additional worker is equal to this wage. Finally, the mass of new hires into each job exactly equals the mass of workers who exogenously separated from the job in the previous period. The resulting solution, which we refer to as the Shapiro-Stiglitz solution and denote with the superscript ss, is described in the following Lemma.

LEMMA 1. A firm that maximizes its profits separately over the two tasks chooses wages $w_i^{ss} = c_i + (1 - (1 - d_i)\,\delta)\,/q_i\,(1 - d_i)\,\delta c_i$ that provide rents $v_i^{ss} = c_i/\delta q_i\,(1 - d_i)$ to each worker performing task $i = 1, 2$. Further, such a firm hires $H_i^{ss} = (1 - d_i)\,N_i^{ss}$ workers, where $\partial f\left(N_i^{ss}, N_{-i}^{ss}\right)/\partial N_i = w_i^{ss} > c_i$.

Lemma 1 reproduces the following key observations from Shapiro and Stiglitz. First, the firm must give the worker rents to provide incentives for the worker to exert effort. Second, the size of the rents required to provide incentives, and hence the wage level, increases in the turnover rate d_i and decreases in the firm's ability to monitor the worker q_i. Third, the firm optimally chooses an employment level for each task that is smaller than the socially optimal level, which

would satisfy $\partial f/\partial N_i = c_i$. Moreover, the gap between the firm's employment-level choice and the socially optimal level is greater for jobs that require higher rents to provide incentives. To facilitate our discussion, define

$$R_i = \frac{c_i}{(1-d_i)\,\delta q_i}$$

as the **Shapiro-Stiglitz rents** associated with task i. We assume throughout that $R_2 > R_1$, so that in the efficiency-wage benchmark, more rents are provided to workers in task 2 than in task 1.

4 Managing Careers

In the efficiency-wage benchmark, the firm chooses only the mass of workers employed in each task and the wage paid to each of these workers. But the number of slots and the wage policy are not the only instruments the firm has at its disposal in this setting. More generally, hiring into each task need not be chosen to exactly offset the exogenous turnover of workers in that task, workers assigned to task i in period t can be assigned to task $j \neq i$ in period $t+1$, and the firm need not retain all workers in task i who do not depart exogenously. In this section, we demonstrate that if the firm utilizes this richer set of instruments, then the firm always performs better than it does if it treats the jobs independently and pays efficiency wages, and we characterize the firm's optimal choices.

4.1 Preliminaries

The firm chooses the wages, hiring, firing, and internal mobility decisions jointly to maximize its steady-state profits

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

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

Promise-Keeping Constraints. Assuming that workers always exert effort, their payoffs must satisfy

the following equations

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

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

Individual Rationality Constraints. To ensure that the workers prefer working for the firm rather than taking their outside options, they must receive a greater payoff from doing so. That is,

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

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

Incentive Compatibility Constraints. Workers have incentives to exert effort if the following constraints are satisfied:

$$\begin{aligned} w_1 - c_1 + (1 - d_1)\,\delta\,(p_{11}v_1 + p_{12}v_2) &\geq w_1 + (1 - q_1)\,(1 - d_1)\,\delta\,(p_{11}v_1 + p_{12}v_2)\,; \\ w_2 - c_2 + (1 - d_2)\,\delta\,(p_{21}v_1 + p_{22}v_2) &\geq w_2 + (1 - q_2)\,(1 - d_2)\,\delta\,(p_{21}v_1 + p_{22}v_2)\,, \end{aligned}$$

where we use the fact that if the worker leaves the firm, he receives a payoff of 0. For notational convenience, we rewrite these constraints as follows:

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

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

where R_i is the Shapiro-Stiglitz efficiency rent associated with task $i = 1, 2$. It will sometimes be useful to denote the excess rents offered in task i by $\Delta_i \equiv p_{i1}v_1 + p_{i2}v_2 - R_i$.

Flow Constraints. In the steady state, the number of workers in a particular task must remain constant. Given promotion, job-security, and recruiting policies, the following constraints ensure that the mass of workers flowing into each task is equal to the mass of workers flowing out of that task:

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

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

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

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

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

We solve the firm's problem in two steps. First, we fix the mass of workers performing each task, and we solve for the firm's cost-minimizing levels of p_{ij}, H_i, and v_i. In the second step, we allow the firm to optimize over N_1 and N_2. Throughout, we refer to the ratio N_2/N_1 as the firm's **span** and N_1 as the firm's **size**. The vector $H = [H_i]_i$ is the firm's **recruiting policy**, and the rent vector $v = [v_i]_i$ is the firm's **wage policy**. The values $1 - p_{i1} - p_{i2}$ represent the probability that the firm asks a productive worker in task i to leave the firm, so the matrix $P = [p_{ij}]_{ij}$ represents both the firm's **promotion policy** as well as its **job-security policy**. If $1 - p_{i1} - p_{i2} = 0$, we say that task i has **full job security**; that is, workers in task i depart the firm only for exogenous reasons. Finally, we define $M_i = (1 - d_i)\,N_i$ to be the **incumbent workers in task** i.

4.2 Internal Organization

Given the span and size of the firm, the firm's steady-state wage bill is $N_1 w_1 + N_2 w_2$. In this section, we solve for the firm's cost-minimizing choices of promotion, hiring, wage, and job-security policies. For reasons that will soon become clear, we refer to task 1 as the **bottom job** and task 2 as the **top job**. It is useful first to rewrite the firm's wage bill in terms of the payoffs that are promised to workers in task i rather than the wages.

To promise lifetime payoffs v_i, the firm must pay wages w_i in each task to cover the costs the worker incurs and to service the promised payoffs. These payoffs are serviced both through today's

wages in excess of worker costs and through future payoffs, accounting for the fact that workers may move about the organization. Since the flow constraint implies that workers who enter job i from somewhere in the organization must equal those who work in job i today in excess of those hired from outside today, the total future payoffs promised to today's workers can be expressed as the total payoffs promised to all those who work at the firm tomorrow in excess of those who are hired tomorrow. Wages in excess of effort costs can therefore be split into two components: the first component is allocated toward promises that are made to new workers, and the second component services existing promises to incumbent workers. The rents promised to incumbent workers must be weakly greater than those required to maintain their incentives to work today. That is, substituting the flow, promise-keeping, and incentive-compatibility constraints, the total wage bill net of effort costs is given by

$$H_1 v_1 + H_2 v_2 + (1-\delta)\left(M_1\left(R_1+\Delta_1\right) + M_2\left(R_2+\Delta_2\right)\right). \qquad \text{(Wage Bill)}$$

The firm seeks to minimize the wage bill subject to $(IR-1)$, $(IR-2)$, $(IC-1)$, $(IC-2)$, $(FL-1)$, and $(FL-2)$. An **internal labor market** is a vector (P, H, v) of promotion policies, job-security policies, hiring policies, and wage policies. An internal labor market (P, H, v) is **feasible** if it satisfies the promise-keeping, incentive-compatibility, individual-rationality, and flow constraints. An internal labor market that results in lower costs than any other feasible internal labor market is referred to as an **optimal internal labor market.**

Throughout, we will assume that in an optimal internal labor market, the rents provided in the top job exceed those provided in the bottom job (i.e., $v_2^* > v_1^*$) and the firm never demotes workers at the top (i.e., $p_{21}^* = 0$), both of which we will later show are indeed features of an optimal internal labor market.

Assume that $N_2 d_2 < (1-d_1) N_1$, so that the number of voluntary departures from the top job is not large enough to accommodate promoting all workers at the bottom of the organization. It must be the case that the firm either demotes or releases workers at the top or does not promote all workers at the bottom. When this is the case, the firm will never hire at the top (i.e., $H_2^* = 0$).

OBSERVATION 1. *In an optimal internal labor market, the firm never hires into the top job unless there are insufficient incumbent workers to fill the slots required for the top job.*

To see why this must be the case, note that because $v_2^* > v_1^*$, the wage bill would be weakly lower if the firm were to reduce the mass of hires at the top and replace them with the same number of hires at the bottom. In conjunction with this shift in hiring policy, the firm must either increase the probability of promotion at the bottom or increase the probability of retention at the top in order to satisfy the flow constraints. One of these changes will always be possible, and either of these changes maintains one of the incentive-compatibility constraints while strengthening the other. Such a change would therefore be feasible and profit-increasing, which contradicts the idea that hiring at the top is optimal.

The next observation is that the incentive-compatibility constraint is always binding for workers at the top job (i.e., $\Delta_2^* = 0$), for if it were not, then the firm could always reduce the wages at the top and maintain both incentive-compatibility constraints. If $p_{12}^* < 1$, then this decrease in wages at the top would need to be accompanied by an increase in p_{12}, perhaps with an accompanying decrease in p_{22}. If $p_{12}^* = 1$, then it must be the case that $p_{22}^* < 1$, in which case since the top job requires more rents than the bottom job, if $(IC-2)$ holds, then $(IC-1)$ holds.

In the latter case, where $p_{22}^* < 1$, it must be the case that both $(IC-1)$ and $(IC-2)$ are binding. Otherwise, the firm could decrease the probability of promotion p_{11}^* and increase the job security at the top, p_{22}^*, which would in turn allow the firm to decrease the wages at the top while still maintaining incentives to exert effort.

OBSERVATION 2. *In an optimal internal labor market, the incentive-compatibility constraint is always binding for workers at the top of the organization. If the firm does not offer full job security at the top, then the incentive-compatibility constraint is also binding for workers at the bottom.*

Since the firm does not hire at the top ($H_2^* = 0$), and the incentive-compatibility constraint is binding at the top ($\Delta_2^* = 0$), the steady-state wage bill can be written as the sum of a constant plus $H_1 v_1 + (1-\delta)\,\Delta_1$. Reducing v_1 reduces the firm's wage bill. When the incentive-compatibility constraint at the bottom is not binding, v_1 can be reduced without affecting incentives at the

bottom or at the top, since $p_{21}^* = 0$. When the incentive-compatibility constraint at the bottom is binding, then $\Delta_1^* = 0$ and the wage bill is clearly decreasing in v_1. Either way, in an optimal internal labor market, workers at the bottom receive no rents. That is, $v_1^* = 0$.

If no rents are provided at the bottom, then workers at the top do not place any value on their future within the firm if they are demoted, so the firm can, without loss of generality, fire any worker at the top who does not stay at the top job. That is, $p_{21}^* = 0$ is in fact optimal. This result relies on the implicit assumption that the outside option for workers at the top job is weakly higher than the outside option for workers at the bottom job, a feature which will remain in equilibrium in later sections when we endogenize the workers' outside options. Further, since the incentive-compatibility constraint is binding at the top, it must be that $v_2^* > 0$, which implies that $v_2^* > v_1^*$, so both of the assumptions we made on endogenous variables are in fact satisfied at the optimum.

OBSERVATION 3. *In an optimal internal labor market, workers at the bottom receive no rents. Furthermore, workers are never demoted, and rents are higher at the top than at the bottom.*

Next, in order to understand when workers at the top will be pushed out of the organization (i.e., $p_{22}^* < 1$), 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 efficiency-wage benchmark in which there are no promotions, and each task is associated with full job security and is paid a wage that corresponds to its Shapiro-Stiglitz rents. At the end of any period, $d_2 N_2$ 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 are a mass $(1 - d_1) N_1$ of incumbent workers at the bottom who must be promised rents R_1 to exert effort. We say that **rents are not scarce at the top** if the workers at the bottom job could be fully motivated to exert effort simply by the promise of being promoted to the top job to fill the vacancies created by the recent departures; that is, if $d_2 N_2 R_2 > (1 - d_1) N_1 R_1$. If this condition is not satisfied, we say that **rents are scarce at the top**.

When rents are not scarce at the top, there is no need to push workers at the top out of the organization (i.e., $p_{22}^* = 1$), since promoting workers to fill the slots opened by exogenous turnover at the top is sufficient for motivating the workers at the bottom. In this case, the incentive-

compatibility constraint for the top job implies the rents provided at the top are equal to the Shapiro-Stiglitz rents (i.e., $v_2^* = R_2$). When rents are scarce at the top, in order to ensure zero rents at the bottom, there must be more promotions offered to workers at the bottom than there are slots opened up by exogenous turnover at the top. That is, the firm must endogenously create turnover at the top and will therefore choose $p_{22}^* < 1$. In turn, the incentive-compatibility constraint for the top job implies that the rents provided at the top are higher than the Shapiro-Stiglitz rents (i.e., $v_2^* > R_2$).

OBSERVATION 4. *In an optimal internal labor market, workers at the top receive full job security if and only if rents are scarce at the top. If rents are scarce at the top, workers at the top receive wages in excess of Shapiro-Stiglitz wages. Otherwise, they receive Shapiro-Stiglitz wages.*

Finally, throughout, we have assumed that $d_2 N_2 < (1 - d_1) N_1$. When the opposite is true, there are not enough incumbent workers in the organization to fill the vacancies that are generated in the top job in every period, so the optimal internal labor market necessarily promotes all workers at the bottom and offers full job security to workers at the top, and the firm hires workers into the top to close the gap between the incumbent workers and N_2 (i.e., $H_2^* = N_2 - M_1 - M_2$). The corresponding analog of Observations $(1) - (4)$ continue to hold in this case. These results are summarized in Proposition 1.

PROPOSITION 1. *In an optimal internal labor market, hiring occurs only in the bottom job, unless there are insufficient incumbent workers to fill the slots required for the top job. Workers in the top job are never demoted. Further, they receive wages in excess of Shapiro-Stiglitz wages if and only if they do not receive full job security, which occurs if and only if rents are not scarce at the top. Workers in the bottom job receive full job security and are promoted to fill all empty slots in the top job.*

Proposition 1 has a number of implications for the firm's hiring and promotion policies. First, there is a single port of entry into the firm: the firm hires into the bottom job but not the top job. Second, there is a well-defined career path for workers in the firm. Workers in the bottom

job either stay at the bottom job or are promoted to the top job. Workers at the top job either stay at the top job or are asked to leave the firm–such workers are never demoted. These patterns reflect several of the key features of internal labor markets reported by Doeringer and Piore (1971). Proposition 1 therefore provides a cost-minimization rationale for the emergence of internal labor markets, and their structure.

To see why firms prefer to start workers at the bottom job, notice that hiring a new worker directly into the top job requires that the firm provide positive rents to an outside worker. In contrast, promoting a worker from within allows the firm instead to give these rents to a worker who is already inside the firm. By filling the top job with incumbent workers, the firm is able to use the rents associated with the top job both to provide incentives for the workers at the top job and to motivate the workers at the bottom job. In other words, promotions from within allow the rents at the top to be reused, a result that is reminiscent of Board (2011).

When rents are scarce at the top, the prospect of being promoted to fill one of the slots generated by exogenous turnover is not enough to provide incentives for workers at the bottom. There are two potential ways the firm could fill this gap. First, the firm could simply increase wages at the bottom, which would increase rents while holding the promotion probability constant. Second, the firm could increase promotion opportunities by freeing up slots at the top through endogenous turnover. But lowering the retention probability of the top workers implies that the wages at the top must increase in order to maintain incentives for the workers at the top.

Surprisingly, when rents are scarce at the top, promoting turnover at the top is optimal. Workers at the top may therefore be fired with positive probability, even if they have not been caught shirking, and even if firing workers at the top requires increasing their wages. This result follows partly from the logic of backloaded payment schemes: workers at the bottom value the possibility of achieving high rents at the top, and the firm can extract these rents from such workers through a low wage at the bottom job. However, such a scheme must be carried out in a cost-minimizing way that maintains incentives for workers at the bottom as well as the top. Consequently, wages are increased at the top job in conjunction with a decrease in the job security at the top.

Another implication is that optimal internal labor markets increase the wage gap between the

top job and the bottom job, relative to the efficiency-wage benchmark. Workers in the bottom job receive wages below the Shapiro-Stiglitz wage for that job, and workers at the top receive a wage that is equal to the Shapiro-Stiglitz wage when rents are not scarce at the top, and they receive a wage that exceeds the Shapiro-Stiglitz wage when rents are scarce at the top. The wage gap is larger when there are more workers at the bottom or fewer workers at the top. It is well-known that wage growth is higher for employed workers than self-employed ones. Notice that while the idea that wages should be backloaded is well-understood, our model points out that this higher wage carries its own risks: in particular, when additional turnover is needed to provide incentives at the bottom, Proposition 1 shows that a higher wage is associated with higher employment risk.

4.3 Optimal Firm Size and Span

In the previous section, we took the firm's size and span as given and solved for the cost-minimizing internal labor market. We now turn to endogenizing the size and span of the firm. It will be useful to write the minimized total wage bill as follows:

$$W(N_1, N_2) = \min_{P,H,w} w_1 N_1 + w_2 N_2 = \tilde{c}_1 N_1 + \tilde{c}_2 N_2,$$

where the coefficients $\tilde{c}_1$ and $\tilde{c}_2$ depend on whether the firm optimally chooses to operate in the case in which rents are scarce at the top (i.e., $d_2 N_2 R_2 < (1 - d_1) N_1 R_1$). When rents are not scarce at the top, the incentive-compatibility constraint for workers at the bottom is not binding, so the cost of adding another position at the bottom is determined only by the cost of effort for the bottom job. The cost of adding a position at the top is determined by the Shapiro-Stiglitz wage that must be paid for the top job, so if we let $\kappa = \frac{1-d_1}{d_2} \frac{R_1}{R_2}$, when $N_2 > \kappa N_1$,

$$\begin{aligned} \tilde{c}_1 &= c_1 \\ \tilde{c}_2 &= \left(1 + \frac{1/(1-d_2) - \delta}{\delta q_2}\right) c_2. \end{aligned}$$

When rents are scarce at the top, the incentive-compatibility constraint is binding for workers at both levels of the organization. Adding another position at the bottom and adjusting the optimal

internal labor market requires that workers at the top be pushed out with a higher probability, which increases the wage they must be paid in equilibrium, so the cost of adding another position at the bottom is greater than c_1. Relatedly, adding another position at the top allows the firm to decrease the probability that workers at the top are pushed out and therefore allows the firm to reduce the wages that must be paid at the top. Therefore, the cost of adding another position at the top is less than the Shapiro-Stiglitz wage for the top position. When $N_2 < \kappa N_1$, we have that

$$\begin{aligned} \tilde{c}_1 &= \left(1 + \frac{1-\delta}{\delta q_1}\right) c_1 \\ \tilde{c}_2 &= \left(1 + \frac{1-\delta}{\delta q_2}\right) c_2. \end{aligned}$$

Figure 1 below combines these two cases. The resulting isocost curve is kinked around the $N_2 = \kappa N_1$ ray. The downward-sloping dotted line represents the isocost curve (corresponding to the same cost level) that would result if effort were contractible. It is immediate that the level of production will be distorted, since for a given cost level, the maximum quantity that can be produced at that cost is lower. We will return to the question of optimal firm size shortly, but for now, we focus on the optimal span $\left(\frac{N_2^*}{N_1^*}\right)$. If we assume that the production function is homothetic (e.g., $f(N_1, N_2) = \left(N_1^{\alpha} N_2^{1-\alpha}\right)^{\sigma}$ with $\sigma \leq 1$), then if effort were contractible, efficient production would occur along the ray $N_2 = \frac{1-\alpha}{\alpha} \frac{c_1}{c_2} N_1$.

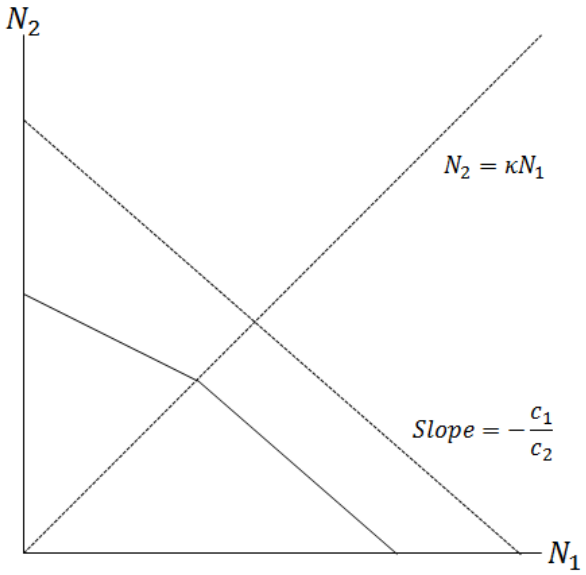

Figure 1

Figure 1 plots an isocost curve for the model in which effort is contractible (downward-sloping dotted line) and the corresponding isocost curve, for the same cost level, for the optimal internal labor market (kinked solid curve).

In contrast, if effort is not contractible, under the optimal internal labor market, production takes the following form: there exists two cutoff values $\alpha_1 < \alpha_2$ such that if $0 \leq \alpha < \alpha_1$, production occurs along the ray $N_2 = \frac{1-\alpha}{\alpha} \frac{1}{1+\frac{1/(1-d_2)-\delta}{\delta q_2}} \frac{c_1}{c_2} N_1$, which is in the region in which rents are not scarce at the top. Production occurs at a lower level and is distorted towards a higher mix of N_1 positions relative to N_2 positions. If $\alpha_2 < \alpha \leq 1$, then production occurs along the ray $N_2 = \frac{1-\alpha}{\alpha} \frac{1+\frac{1-\delta}{\delta q_1}}{1+\frac{1-\delta}{\delta q_2}} \frac{c_1}{c_2} N_1$, which is in the region in which rents are scarce at the top. Production occurs at a lower level and may be distorted in either direction relative to the contractible effort case, depending on whether q_1 is greater than or less than q_2. Finally, if $\alpha_1 \leq \alpha \leq \alpha_2$, production occurs at the kink. These cases are shown in Figure 2.

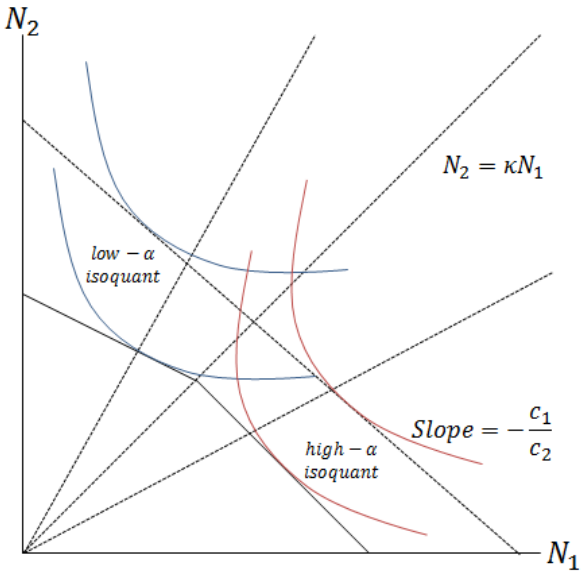

Figure 2

Figure 2 plots two sets of isoquants, corresponding to low- and high-α production functions. The blue isoquants correspond to low-α production: the outer curve represents optimal production under contractible effort, and the inner curve represents optimal production under the optimal internal labor market. The red isoquants similarly correspond to high-α production.

For non-homothetic production functions, the production expansion path is non-linear, but for homothetic production functions, the span $\left(\frac{N_2^*}{N_1^*}\right)$ is pinned down independently of the size (N_1^*) of the firm. Given $0 \leq \alpha < \alpha_1$, so that rents are not scarce at the top, we have:

$$\begin{aligned} N_2 &= \frac{1-\alpha}{\alpha} \frac{1}{1+\frac{1/(1-d_2)-\delta}{\delta q_2}} \frac{c_1}{c_2} N_1 \\ y\left(N_1, N_2^*\left(N_1\right)\right) &= N_1^\sigma \left(\frac{1-\alpha}{\alpha} \frac{1}{1+\frac{1/(1-d_2)-\delta}{\delta q_2}} \frac{c_1}{c_2} \right)^{(1-\alpha)\sigma} \\ W\left(N_1, N_2^*\left(N_1\right)\right) &= \frac{c_1}{\alpha} N_1 \end{aligned}$$

Given $\alpha_1 \leq \alpha \leq \alpha_2$,

$$\begin{aligned} N_2 &= \frac{1-d_1}{d_2}\frac{R_1}{R_2}N_1 \\ y\left(N_1, N_2^*\left(N_1\right)\right) &= N_1^{\sigma}\left(\frac{1-d_2}{d_2}\frac{c_1}{c_2}\frac{q_2}{q_1}\right)^{(1-\alpha)\sigma} \\ W\left(N_1, N_2^*\left(N_1\right)\right) &= c_1N_1 + c_2N_2 + (1-\delta)N_2v_2 \\ &= \left(1 + \frac{\delta\left(1-d_2\right)q_2 + 1 - \delta}{\delta d_2 q_1}\right)c_1N_1 \end{aligned}$$

and given $\alpha_2 < \alpha \leq 1$, so that rents are scarce at the top

$$\begin{aligned} N_2 &= \frac{1-\alpha}{\alpha}\frac{1+\frac{1-\delta}{\delta q_1}}{1+\frac{1-\delta}{\delta q_2}}\frac{c_1}{c_2}N_1 \\ y\left(N_1, N_2^*\left(N_1\right)\right) &= N_1^{\sigma}\left(\frac{1-\alpha}{\alpha}\frac{1+\frac{1-\delta}{\delta q_1}}{1+\frac{1-\delta}{\delta q_2}}\frac{c_1}{c_2}\right)^{(1-\alpha)\sigma} \\ W\left(N_1, N_2^*\left(N_1\right)\right) &= \left(1 + \frac{1-\delta}{\delta q_1}\right)\frac{c_1}{\alpha}N_1. \end{aligned}$$

5 Effects of Labor Market Policies

In this section, we use our model to study how three common public policies affect careers in organizations. These policies are progressive taxes that increase the cost of compensating workers in the top job relative to those in the bottom job, employment-protection policies that increase the cost of firing workers without cause, and minimum wages. Our model will allow us to describe how these policies affect the firm's employment level, span, wage levels, and promotion probabilities.

Throughout this section, we will assume that, in both jobs, shirking is always detected (i.e., $q_1 = q_2 = 1$). In the first two applications, our analysis focuses primarily on the case in which the incentives of the bottom workers is important. While progressive taxes and firing restrictions directly affect workers at the top of the organization, when rents are scarce at the top, we show that these policies can indirectly affect workers at the bottom as well, as they affect the optimal internal labor market. In our minimum wage application, we show that the employment level can

increase with the minimum wage.

5.1 Progressive Taxes

Consider the following progressive tax scheme. Wages are taxed at rate 0 until a threshold w^*, after which they are taxed at a rate $t > 0$. The progressive tax scheme can be summarized by the vector (t, w^*). Under tax scheme (t, w^*), if the firm pays a wage w, the worker receives

$$w' = \begin{cases} w & \text{if } w \leq w^*; \\ w^* + (1-t)(w - w^*) & \text{if } w > w^*. \end{cases}$$

Many of the main features of the baseline model are preserved under any progressive tax scheme. As before, all new workers enter at the bottom job, and existing workers are never demoted. Clearly, when the wage associated with the top job in the baseline model is below w^*, a progressive tax scheme (t, w^*) has no effect on the firm's internal labor market. However, when the wage of the top job in the baseline model exceeds w^*, raising the wage of the top job becomes relatively more costly for the firm. As a result, the firm will adjust its wage and promotion policies depending on the tax rate. The next proposition describes the firm's wage and promotion policies, for a fixed number of positions N_1, N_2 in the firm.

The following tax rate, which we denote as t^* will serve as an important cutoff for our analysis:

$$t^* \equiv d_1 + \frac{N_2}{N_1} \frac{\delta w^* - c_2}{\delta w^*}. \qquad \text{(Cutoff Tax Rate)}$$

We show in the following proposition that when the tax rate at the top is below this threshold, the firm adjusts its wage policy to account for for the higher cost of compensating workers at the top, but it does not adjust any of its other policies. The results are quite different for tax rates exceeding this threshold, however.

PROPOSITION 2. *Suppose rents are scarce at the top, and* $N_1 > N_2 (\delta w^* - c_2) / c_1$. *Then:*

(i): *If* $t < t^*$,

$$\begin{aligned} w_1 &= 0; \quad w_2 = \frac{1}{(1-t)\,N_2}\left((1-d_1)\,N_1R_1 + (1-d_2)\,N_2R_2 - tN_2w^*\right); \\ p_{12} &= \frac{N_2R_1}{(1-d_1)\,N_1R_1 + (1-d_2)\,N_2R_2}; \quad p_{22} = \frac{N_2R_2}{(1-d_1)\,N_1R_1 + (1-d_2)\,N_2R_2}. \end{aligned}$$

(ii): *If* $t > t^*$,

$$\begin{aligned} w_1 &= \frac{c_1N_1 - (\delta w^* - c_2)\,N_2}{\delta\,(1-d_1)\,w^*N_1 - (\delta w^* - c_2)\,N_2}w^*, \quad w_2 = w^*; \\ p_{12} &= \frac{N_2}{(1-d_1)\,N_1}\left(1 - \frac{c_2}{\delta w^*}\right); \quad p_{22} = \frac{c_2}{\delta\,(1-d_2)\,w^*}. \end{aligned}$$

The conditions in Proposition 2 ensure that rents are scarce at the top and that the progressive tax policy matters (i.e., the wage associated with the top job in the baseline model exceeds w^*. This parameter range allows us to focus on the more interesting case in which the incentives of bottom workers are important. Part (i) shows that when the tax rate is smaller than t^*, the wage and promotion policy is very similar to the baseline case. In particular, workers at the bottom receive zero rents and therefore zero wages. Workers at the top also receive the same payoff as in the baseline case, and as a result, the promotion probability (p_{12}) and the retention probability of the top job (p_{22}) remain unchanged. The only difference is that, because of the positive tax rate, the pre-tax wage paid to workers at the top must correspondingly increase to maintain the same post-tax wages.

When the tax rate exceeds t^*, the results differ strikingly. Part (ii) shows that workers at the bottom receive a positive rent, and therefore a positive wage. Workers at the top receive a wage of w^*, which by construction is lower than the wage they receive in the baseline case. In other words, progressive taxes compress the post-tax wage structure significantly when the tax rate exceeds t^*. Consequently, the promotion prospects for workers at the bottom are increased, and workers at the top are granted higher job security. The intuition is straightforward: when the tax rate is sufficiently high, it is too costly for the firm to raise wages at the top above w^* to provide incentives for workers at the bottom. Thus, the firm sets the top job's wage at w^* and raises the

wages of the bottom job to provide incentives for workers at the bottom.

Proposition 2 describes the effects of progressive taxes on the wage and promotion policies of the firm, fixing the number of positions N_1 and N_2. In general, progressive taxes also affect the hierarchical structure of the firm, which in turn also affects wages and promotion probabilities. To understand these effects, we again consider the case in which production is Cobb-Douglas and given by $f(N_1, N_2) = \left(N_1^{\alpha} N_2^{1-\alpha}\right)^{\sigma}$ with $\sigma < 1$. In addition, for tractability, we focus on progressive tax schemes with $t < t^*$. The next corollary summarizes the effect of progressive taxes on the optimal internal labor market.

COROLLARY 1. *Suppose rents are scarce at the top and* $\frac{\alpha}{1-\alpha} > \frac{\delta w^* - c_2}{c_2 - \delta t w^*}$. *In this case, the number of positions in the firm are given by*

$$\begin{aligned} N_1 &= (\delta(1-t))^{\frac{1}{1-\sigma}} \sigma^{\frac{1}{1-\sigma}} \alpha^{\frac{1-(1-\alpha)\sigma}{1-\sigma}} (1-\alpha)^{\frac{(1-\alpha)\sigma}{1-\sigma}} c_1^{\frac{(1-\alpha)\sigma-1}{1-\sigma}} (c_2 - \delta t w^*)^{\frac{-(1-\alpha)\sigma}{1-\sigma}}; \\ N_2 &= (\delta(1-t))^{\frac{1}{1-\sigma}} \sigma^{\frac{1}{1-\sigma}} \alpha^{\frac{\alpha\sigma}{1-\sigma}} (1-\alpha)^{\frac{1-\alpha\sigma}{1-\sigma}} c_1^{\frac{\alpha\sigma}{1-\sigma}} (c_2 - \delta t w^*)^{\frac{\alpha\sigma-1}{1-\sigma}}. \end{aligned}$$

In addition, the following hold.

(i): The span of the firm (N_1/N_2) *is given by*

$$\frac{N_1}{N_2} = \frac{\alpha}{1-\alpha} \frac{c_2 - \delta t w^*}{c_1}.$$

(ii): The promotion probability (p_{12}) *is given by*

$$\frac{(1-\alpha) c_1}{(1-d_1)(c_2 - \alpha \delta t w^*)}.$$

(iii): The wage of the top job (w_2) *is given by*

$$\frac{c_2 - \delta t w^*}{\delta(1-\alpha)(1-t)}.$$

Notice that when the tax rate is $t = 0$, the span, promotion probability, and wages are the same as those in the baseline case. As the tax rate increases, the span decreases, and the promotion

probability increases. The reason is that, as t increases, it becomes more costly to increase the wage (holding constant the level of take-home pay). As a result, the firm adjusts instead by increasing the number of positions at the top relative to the number at the bottom. This change results in a decrease in the span and an increase in the promotion probability.

Additionally, the pre-tax wages at the top increase, but the post-tax wages at the top actually decrease with the tax rate. This follows, because a higher tax rate makes it more costly to motivate the workers at the bottom by the wage increase upon promotion. Instead, the firm adjusts the probability of promotion upward and the wage increase upon promotion downward. This results in a lower post-tax wage for the workers at the top, although they are partially compensated by improved job security (i.e., p_{22} is higher).

Finally, the effect of the tax rate on the number of positions in the firm is more complicated, because there are two forces that go in opposite directions. On the one hand, a higher tax rate makes it more costly to hire workers, which is reflected by the $(1-t)$ term. On the other hand, a higher tax rate causes the firm to reduce the wages at the top. To maintain incentives, the firm resorts to increasing the number of positions at the top in order to increase the promotion prospects for workers at the bottom. This positive effect of the tax rate on the number of positions in the firm is reflected by the $(c_2 - \delta t w^*)^{\frac{\alpha\sigma-1}{1-\sigma}}$ term in N_2. Notice that once the firm increases the number of positions at the top, it increases the marginal productivity of the workers at the bottom. This force also causes the firm to increase the number of positions at the bottom. For the bottom job, the production effect dominates the incentive effects, and the number of workers at the bottom decreases in the tax rate.

5.2 Firing Restrictions

Our second application examines the effects of employment-protection policies that increase the cost of firing workers without cause–that is, firing a worker who has not been caught shirking. We look at an extreme policy in which the cost of firing without cause is infinity. Further, in line with many observed employment-protection policies, the firm cannot force the worker out of the organization by offering an arbitrarily low wage or by demoting the worker.

As in the progressive-taxes application, the main features of internal labor market do not change once the firm is unable to fire workers without cause. Again, there is port of entry into the firm via the bottom job, and there is a well-defined career trajectory within the firm. Obviously, when there are sufficient rents at the top to ensure that no in the baseline model, full job security would be optimal, firing restrictions do not affect the optimal internal labor market. When rents are scarce at the top, however, raising wages at the top without being able to fire workers at the top becomes more costly. This leads the firm to adjust its wage and promotion policies, which in turn affects the firm's hierarchical structure.

The next proposition describes how firing restrictions affect the wage and promotion policies, fixing the number of positions in the firm. As in the progressive-taxes application, there will be two cases that characterize the effects of firing restrictions. The following condition is a necessary condition for when workers at the bottom receive no rents:

$$(1-\delta)\,N_1 + \delta H_1 < (1-\delta)\,\frac{N_1 - H_1}{d_2}, \tag{No-Rent}$$

where $H_1 = d_1 N_1 + d_2 N_2$ is the number of new hires the firm has in each period. We refer to this condition as the **No-Rent Condition**.

PROPOSITION 3. *Suppose rents are scarce at the top.*

(i): *If the No-Rent Condition is satisfied,*

$$\begin{aligned} w_1 &= 0; \quad w_2 = \frac{1-\delta\,(1-d_2)}{d_2}\frac{c_1}{\delta}\frac{N_1}{N_2} + c_2; \\ p_{12} &= \frac{d_2}{1-d_1}\frac{N_2}{N_1}. \end{aligned}$$

(ii): *If the No-Rent Condition is not satisfied,*

$$\begin{aligned} w_1 &= \frac{(1-d_2)\,c_1 N_1 - ((1-d_1)\,N_1 - d_2 N_2)\,c_2}{\delta\,((1-d_1)\,N_1 - d_2 N_2)\,(1-d_2)}, \quad w_2 = \frac{c_2}{\delta\,(1-d_2)}; \\ p_{12} &= \frac{d_2}{1-d_1}\frac{N_2}{N_1}. \end{aligned}$$

Part (i) of Proposition 3 shows that when the No-Rent Condition is satisfied, workers at the bottom again receive a wage of zero, as in the baseline model. Payoffs for workers at the top, however, are higher than in the baseline model (for a fixed N_1 and N_2). The reason is that when the firm cannot fire workers at the top, the upward mobility of workers at the bottom becomes limited. In order to maintain incentives for workers at the bottom, the firm must increase the value of the top job. In fact, the incentive-compatibility constraint for workers at the top is slack in this case, so the workers at the top strictly prefer working to shirking–the extra rents they receive are not used to motivate them, but rather to motivate the workers at the bottom.

Part (ii) shows that when the No-Rent condition is not satisfied, workers at the bottom receive strictly positive rents and wages. The wages at the top are the same as the Shapiro-Stiglitz wages. To supplement the incentives at the bottom, the firm uses the efficiency wage in this case rather than strengthening the promotion incentives, as in Part (i). Notice that the No-Rent Condition is not satisfied only when the turnover rate is sufficiently low. In this case, promotion opportunities are sufficiently small that a large increase in the wage at the top is needed to increase incentive at the bottom. This makes it more cost-effective to use efficiency wages at the bottom to provide incentives.

As in the progressive-taxes application, we next consider the effect of firing restrictions on the number of positions in the firm. Again, we consider the Cobb-Douglas production function and focus on case where the turnover rate is not too low. The next corollary summarizes the effects of firing restrictions.

COROLLARY 2. *Suppose rents are scarce at the top. Then the number of positions in the firm are given by*

$$\begin{aligned} N_1 &= \left(\frac{\delta d_2}{1-\delta\left(1-d_2\right)}\right)^{\frac{1-(1-\alpha)\sigma}{1-\sigma}} \sigma^{\frac{1}{1-\sigma}} \alpha^{\frac{1-(1-\alpha)\sigma}{1-\sigma}} \left(1-\alpha\right)^{\frac{(1-\alpha)\sigma}{1-\sigma}} c_1^{\frac{(1-\alpha)\sigma-1}{1-\sigma}} c_2^{\frac{-(1-\alpha)\sigma}{1-\sigma}}; \\ N_2 &= \left(\frac{\delta d_2}{1-\delta\left(1-d_2\right)}\right)^{\frac{\alpha\sigma}{1-\sigma}} \sigma^{\frac{1}{1-\sigma}} \alpha^{\frac{\alpha\sigma}{1-\sigma}} \left(1-\alpha\right)^{\frac{1-\alpha\sigma}{1-\sigma}} c_1^{\frac{-\alpha\sigma}{1-\sigma}} c_2^{\frac{\alpha\sigma-1}{1-\sigma}}. \end{aligned}$$

In addition, the following hold.

(i): The span of the firm (N_1/N_2) *is given by*

$$\frac{N_1}{N_2} = \frac{\alpha}{1-\alpha}\frac{c_2}{c_1}\frac{\delta d_2}{1-\delta(1-d_2)}.$$

(ii): The promotion probability (p_{12}) *is given by*

$$p_{12} = \frac{1-\alpha}{\alpha}\frac{1-\delta(1-d_2)}{\delta(1-d_1)}\frac{c_1}{c_2}.$$

(iii): The wage of the top job (w_2) *is given by*

$$w_2 = \frac{c_2}{1-\alpha}.$$

Part (i) implies that firing restrictions reduce the span of the firm relative to the baseline model. To see this, recall that the promotion incentives depend both on the wage increase upon promotion as well as the probability of promotion. To maintain incentives at the bottom, the firm can either increase wages at the top or distort the production level by reducing the span, both of which are costly for the firm. Firing restrictions make it more costly for the firm to provide promotion incentives through increasing the wage increase upon, and therefore, the firm optimally adapts to firing restrictions by distorting the production.

Notice that even if the span is smaller, this does not imply that workers' promotion opportunities are greater. There are two forces here relative to the baseline model. On the one hand, reducing the span increases promotion opportunities. On the other hand, firing restrictions imply that the overall turnover (voluntary plus involuntary) rate at the top job is smaller, which in turn reduces the promotion opportunities for those at the bottom. As a result, whether the promotion probability increases or decreases depends on the degree to which the firm reduces its span. Comparing Part (ii) to the benchmark case, the condition for promotion opportunities to decrease is given by $d_2 + \frac{1-\delta}{\delta} < \alpha$. This condition is more likely to be satisfied when α is larger. In this case, the marginal productivity of workers at the top is smaller, so that the value of increasing the number of positions at the top is smaller. This lowers the distortion in the span of the firm, so that the

probability of promotion is smaller when there are firing restrictions.

When the promotion probability is smaller, the value of being promoted must be higher to maintain incentives at the bottom. But even if the value of being promoted is higher, part (iii) implies that the wages at the top job are always smaller, because these workers now have better job security. Raising wages at the top without the corresponding reduction in job security at the top means that the firm must give rents to workers at the top.

5.3 Minimum Wage

Our third application studies the effects of a minimum wage on internal labor markets. In contrast to progressive taxes and firing restrictions, minimum wages have a direct effect on the wages of workers at the bottom. As a result, a minimum wage affects internal labor markets whether or not rents are scarce at the top. Let $\underline{w}$ be the minimum wage. Notice that if $\underline{w} \geq R_1$, the minimal wage alone is enough for motivating the workers in the bottom job, so internal labor markets are no longer valuable. Therefore, we focus on situations in which $\underline{w} < R_1$. The next proposition describes the effects of a minimum wage on the optimal internal labor market, fixing the number of positions in the firm.

PROPOSITION 4. *The following hold.*

(*i*): *If* $N_1 < \frac{d_2}{1-d_1}\frac{R_2-\underline{w}}{R_1-\underline{w}}N_2$, *the minimum wage binds when* $\underline{w} > \frac{(1-d_1)N_1R_1-d_2N_2R_2}{(1-d_1)N_1-d_2N_2}$. *In this case,*

$$\begin{aligned} w_1 &= \underline{w}, \quad w_2 = \frac{c_2}{\delta\left(1-d_2\right)}; \\ p_{12} &= \frac{d_2}{1-d_1}\frac{N_2}{N_1}, \quad p_{22} = 1. \end{aligned}$$

(*ii*): *If* $N_1 \geq \frac{d_2}{1-d_1}\frac{R_2-\underline{w}}{R_1-\underline{w}}N_2$, *the minimum wage binds if and only if* $\underline{w} > 0$. *In this case, workers at the top are kept with probability* p_{22} *and demoted with probability* $1-p_{22}$.

$$
\begin{aligned}
w_1 &= \underline{w}, \quad w_2 = \frac{R_1 N_1 + R_2 M_2 - (M_1 + M_2 - N_2)\underline{w}}{N_2}; \\
p_{12} &= \frac{(R_1 - \underline{w})N_2}{R_1 M_1 + R_2 M_2 - (M_1 + M_2)\underline{w}}, \quad p_{22} = \frac{(R_2 - \underline{w})N_2}{R_1 M_1 + R_2 M_2 - (M_1 + M_2)\underline{w}}.
\end{aligned}
$$

Part (i) describes the case in which rents are not scarce at the top. The wage at top is the same as the Shapiro-Stiglitz wage for the top job, and workers at the top receive full job security. When rents are scarce at the top, part (ii) shows that workers at the top will be demoted rather than fired. The reason for demotion, rather than firing, is that when the minimum wage is positive, rents are positive at the bottom, and therefore the value to an incumbent worker at the top is higher in the bottom position than outside of the firm. In general, if the outside option of workers at the top exceed the rents that the workers at the bottom receive, the firm will use firing rather than demotion.

Notice that, in part (ii), the promotion probability p_{12} is decreasing in $\underline{w}$. Since the minimum wage exceeds the workers' outside option, the minimum wage itself serves as an efficiency wage. A higher minimum wage serves as a stronger efficiency wage, reducing the importance of promotion incentives in motivating workers at the bottom, resulting in a decrease in p_{12}. This logic also explains why job security at the top (p_{22}) increases in the minimum wage: when promotion incentives are less important in motivating workers at the bottom, less distortion is needed at the top. Finally, notice that the incentive constraint for workers at the bottom binds when $\frac{N_1}{N_2} > \frac{d_2}{1-d_1}\frac{R_2-\underline{w}}{R_1-\underline{w}}$. Notice that this ratio is increasing in $\underline{w}$. The reason is again that less rents are need to motivate workers at the bottom when the minimum wage increases.

Next, we consider the effect of a minimum wage on the number of positions in the firm. Unlike the progressive tax or the firing restrictions applications, a minimum wage affects the optimal internal labor market even when rents are not scarce at the top. The next corollary highlights the effect of the minimum wage on the promotion probability for this case.

COROLLARY 3.: *When* $N_1 < \frac{d_2}{1-d_1}\frac{R_2-\underline{w}}{R_1-\underline{w}}N_2$, *the following holds:*

(*i*) *The span of the organization is given by*

$$\frac{N_1}{N_2} = \frac{\delta d_2 R_2}{c_1 - \underline{w}}.$$

(*ii*) *The probability of promotion is given by*

$$p_{12} = \frac{c_1 - \underline{w}}{(1 - d_1)\,\delta R_2}.$$

Perhaps somewhat surprisingly, the firm's span is increasing in the minimum wage, even if rents are not scarce at the top. This effect arises not through the incentive-compatibility constraint of the workers at the bottom, since it is slack, but rather from their participation constraint. In particular, when the minimum wage increases, the participation constraint for new hires is easier to satisfy. As a result, the firm can reduce the probability of promotion by expanding the number of positions at the bottom.

Moreover, when the production function is inflexible in the top job (i.e., if we take N_2 as exogenous) but flexible in the bottom job, it is easy to see that the total employment level of the firm increases in the minimum wage, since the firm's span is increasing in the minimum wage. When the production function depends more flexibly on N_2, a minimum wage makes expanding the number of positions more costly, and the firm may decrease the total number of positions. When the production is Cobb-Douglas, it can be shown that the total number of positions in the firm increases when the effort costs for the bottom job are small relative to those for the top job.

6 Conclusion

This paper studies how the firm optimally designs its organizational structure and human resources policies when the career incentives of its workers are important. Our model builds on the efficiency-wage model of Shapiro and Stiglitz (1984) by allowing the firm to use promotion as an incentive device. A key feature of this model is the flow constraint: the opportunity for promotion is limited by the turnover at the top. Contrary to the Shapiro and Stiglitz model, workers do not receive

rents here, and they are motivated entirely by the promotion incentive. To ensure the promotion incentive is sufficiently strong, the firm sometimes needs to increase the wage of the top job. More importantly, the firm will also increase both the turnover rates the top and create more top positions.

Our model provides a rationale for why internal labor markets are useful: the firm makes more profit by linking the jobs together. Our model also gives rise to a number of well-documented features of the internal labor market: there is a port of entry for new workers and a well-defined career trajectory for existing workers. Moreover, our model sheds light on why firms may benefit from actively managing, and sometimes, increasing, its turnovers. An important tool for increasing turnovers is the use of mandatory retirement policy. Cappelli (20xx) quotes T. V. Houser, vice president of merchandising at Sears, that "retirement was mandatory "entirely to keep the lines of advancement open.""

The simplicity of the model allows us to study the effects of a number of labor market policies on the organization of the internal labor markets. We show that a progressive-tax policy that directly affects the top workers have indirect effects on bottom workers. Fewer workers are hired at the bottom although the existing workers are more likely to be promoted. We also show that by banning the firms from firing its workers, not only the wages of the top job becomes lower and the employment level goes down, but also the span of the hierarchy becomes less flat. Finally, we show that a minimum wage policy can both reduce and increase the employment level, depending on the parameter range.

There are a number of future directions for this model. Currently, all workers are homogenous. By incorporating ability heterogeneity into the model and allowing the ability to be learnt over time, we can use the model to address additional types human resources policies. For example, a direct application of the model is that once internal labor market is important, the firm may prefer promoting from within to hiring from outside, i.e. there can be an insider-bias. In addition, the model can potentially address the wage-seniority puzzle by Medoff and Abraham (1982): workers on a job longer receive higher wages (without performing better) because the chances for promotion for these workers are smaller, so to motivate them, a higher efficiency-wage is called for. Finally,

since our model focuses on managing turnover for promotion purposes, allowing for heterogeneous workers can help shed light on human resources policies (such as up-or-out) in professional industries in which both the incentive and selection of workers are important.

Another direction for future research is to consider firms with multiple levels of hierarchy. This allows us to study not only the turnover policies at the top jobs but also the turnovers in the middle-rank. In addition, multiple levels of hierarchy enable us to make prediction about how wages, promotion probabilities, and spans change at different levels of hierarchy. To mechanism in our model suggests that wage can be convex in hierarchy levels: a larger wage increase at higher level provides incentives for more workers below, c.f. Rosen (1986). Studying the promotion probabilities and spans for multi-layer hierarchies may allow one to uncover further patterns on careers within organization and therefore better understand the effects of labor market policies on internal labor markets.

Third, the outside options of the workers are currently left as exogenous. In human-capital-intensive industries, firms often affect the outside options of their workers by engaging in training in general human capital. Such training is often considered a puzzle because they raise the outside options of the workers; see Becker (1975). In general, the conventional wisdom within economics suggests that a higher outside options of the workers hurt the firms, but in practice firms often take effort to boost the outside options of their workers. In particular, the consulting firms worry about whether their workers can find good employment elsewhere, and one reason appears to be related to creating promotion opportunities. A McKinsey insider comments that "if international companies stopped recruiting former McKinsey staff, it could clog the "up or out" refining process." In our context, if training increases the outside option of the top job more than that for the bottom job, it makes promotion more valuable and can help reduce distortions in the span of hierarchy and wages. This provides a justification for why firms would like to provide general training, which is one way to increase the worker's outside options. More importantly, extending the model to allow for training could shed light on how training policies interacts with the organizational structure of the firm.

Fourth, our model has ruled out formal contracts as a way to motivate the workers. More

generally, firms use a mix of formal contracts and career-based incentives to motivate their workers. For example, one way to increase the turnover rate at the top is to use of buy-out policies. By committing to a severance pay to the top workers, the firms increase the value of promotion, and this makes it easier to motivate the bottom workers. Of course, the effectiveness of these formal contracts depends on the qualities of legal institutions, and when the formal contracts are harder to enforce, the firms reply more on career-based incentives. This model therefore allows us to explore how the availability of formal contracts affects the organization of the internal labor markets and the careers of workers.

Finally, this model considers firms in steady states: the size of the firm and its hierarchy does not change over time. One future research direction is to study how the human resources policies vary for high-growth and more stable more firms. It appears natural that firms with higher growth rate can rely more on promotion incentives to motivate their workers. At some point, however, high-growth firms become more mature and their growth slows down. Understanding how firms will change their human resources policies when the promotion opportunities shrink is a fascinating theoretical question with important practical implications.

7 References

References

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

[2] Beaudry, P., and R. DiNardo, "The Effect of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence from Micro Data," *Journal of Political Economy*, 99 (1991), 665-688.

[3] Becker, G., Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, 2nd Edition. Chicago and London: University of Chicago Press, (1975).

[4] Bernhardt, D. "Strategic Promotion and Compensation," *Review of Economic Studies* 62 (1995), 315-339.

[5] Board, S. "Relational Contracts and the Value of Loyalty," American Economic Review 101(7) (2011), pp. 3349–67.

[6] Cappelli, P., Talent on Demand. Harvard Business Press, (2008).

[7] Chiappori, A, Salanie B, and J. Valentin "Early Starters Versus Late Beginners," *Journal of Political Economy*, 107 (1999), 731-760.

[8] Demougin, D., and A. Siow, "Careers in Ongoing Hierarchies," *American Economic Review*, 84 (1994), 1261-1277.

[9] Dickens William, Kevin Lang, Larry Katz, and Larry Summers (1990), "Why Do Firms Monitor Workers?" in Yoram Weiss and Gideon Fishelson, eds., Advances in the Theory and Measurement of Unemployment (London: MacMillan), pp. 159-71.

[10] Doeringer, Peter and Michael Piore (1971). *Internal Labor Markets and Manpower Analysis.* Lexington, MA: Heath Lexington Books.

[11] Fuchs, William (2007), "Contracting with Repeated Moral Hazard and Private Evaluations", American Economic Review, 97(4); pp. 1432-1448.

[12] Gibbons, R., and M. Waldman, "A Theory of Wage and Promotion Dynamics inside Firms," *Quarterly Journal of Economics*, 114 (1999), 1321-58.

[13] Gibbons, R., and M. Waldman, "Careers in Organizations: Theory and Evidence," Chapter 36 in Volume 3B of O. Ashenfelter and D. Card (eds.), Handbook of Labor Economics, North Holland, (1999).

[14] Hall, Robert (1982), "The Importance of Lifetime Jobs in the U.S. Economy," American Economic Review, 72 (4), pp. 716-724.

[15] Harris, M., and B. Holmstrom, "A Theory of Wage Dynamics," *Review of Economic Studies* 72 (1982) 315-333.

[16] MacDonald, G., "A Market Equilibrium theory of Job Assignment and Sequential Accumulation of Information," *American Economic Review*, 72 (1982) 1038-1055.

[17] MacLeod, B., and J. Malcomson, "Reputation and Hierarchy in Dynamic Models of Employment," *Journal of Political Economy*, 96 (1988) 832-854.

[18] Mincer, J., Schooling, Experiences, and Earnings, New York: Columbia University for National Bureau of Economic Research, (1974).

[19] Neal, D., and S. Rosen., "Theories of the Distribution of Earnings," in Handbook of Income Distributions, Vol. 1, North Holland (2000).

[20] Sattinger, M., "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 31 (1993), 831-80.

[21] Waldman, Michael. "Job Assignments, Signalling, and Efficiency " *RAND Journal of Economics,* Vol. 15, No. 2 (Summer, 1984), pp. 255-267.