

Federal Reserve Bank of Minneapolis
Research Department Staff Report 482

June 2013

Job Matching Within and Across Firms*

Elena Pastorino

Federal Reserve Bank of Minneapolis
and University of Minnesota

ABSTRACT

In order to analyze careers both within and across firms, this paper proposes a matching model of the labor market that extends existing models of job assignment and learning about workers' abilities. The model accounts for worker mobility across jobs and firms, for varying degrees of generality of ability, and for the possibility that firms affect the information they acquire about workers through job assignment. I characterize equilibrium assignment and wages, and show how, depending on how abilities and jobs are distributed across firms, equilibrium gives rise to widely varying patterns of job mobility within firms and turnover across firms, even if matching would be perfectly assortative in the absence of uncertainty. The implied job and wage dynamics display features that are consistent with a broad set of empirical findings on careers in firms and the labor market. In particular, workers can experience gradual promotions and wage increases following successful performance but few or no demotions when employed by the same firm. The model also produces turnover across firms and occupations after both successful and unsuccessful experiences, leading to wage increases or decreases following a firm or occupation change. Overall, the results in this paper provide a unified framework in which to interpret the dynamics of jobs and wages in firms and the labor market.

Keywords: Learning; Matching; Careers in firms; Turnover

JEL Classification: J24, J31, J62

*I have benefited from numerous conversations with Jan Eeckhout, Patrick Kehoe, and Ken Wolpin. I thank Enoch Hill for excellent research assistance and Joan Gieseke for invaluable editorial assistance. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Over their careers, workers typically move through a hierarchy of jobs within a given firm as well as experience turnover between firms. These two types of mobility have been analyzed extensively, albeit separately. A model that integrates both types of job mobility would provide a useful framework in which to interpret the patterns of careers typically observed in the data. In this paper, I develop such a unified model under uncertainty about workers' ability and characterize equilibrium assignment for varying degrees of generality of ability across jobs and firms. I also explore the impact of differences in the speed of learning across jobs on crucial dimensions of the dynamics of jobs and wages. I show that the model's implications are consistent with numerous stylized characteristics of careers, cast new light on some common findings, and naturally explain features of careers previously thought difficult to reconcile with the presence of uncertainty and learning.

I propose an equilibrium model of learning in the labor market in which career paths reflect a gradual process of information acquisition about workers' suitability for the various jobs available in the economy. In this framework a career is the outcome of a two-way feedback between job assignment and information acquisition. As information is acquired, workers optimally advance through jobs both within and across firms. At the same time, workers select into jobs, in part, because of the information acquisition possibilities that employment offers.

The model integrates and extends two leading approaches to the study of careers. Specifically, I extend the approach to the analysis of careers *within* firms developed by Gibbons and Waldman (Gibbons and Waldman (1999, 2006)). Gibbons and Waldman propose an integrated framework of job matching within firms that combines learning and human capital acquisition and qualitatively explains a broad set of empirical findings about job and wage mobility in firms. Their focus is on an environment in which the technologies of all firms are identical, so they abstract from worker turnover. Differently from Gibbons and Waldman, I allow firms to be heterogeneous in their technologies and examine the implications of firm and worker heterogeneity for the patterns of job mobility (and wages) within and across firms.

I nest in this framework and extend the classic approach to worker turnover *across* firms developed by Jovanovic (1979). Jovanovic considers a model of job matching across firms, in which each firm has only one job and workers gradually learn about the quality of their match with a given firm based on their output. By construction, this model is silent about the patterns of careers within firms. Differently from Jovanovic, I allow multiple jobs to exist within firms. Moreover, I explore how different degrees of firm specificity of ability, ranging from purely specific to jobs and firms to purely general across both, lead to different predictions about turnover, both across jobs within a firm and across firms.

Formally, the model is a competitive economy in which a worker's ability, either high or low, is initially unknown to all agents, including that worker. A worker's state is then the

current prior that the worker's ability is high. Firms operate technologies that are described by a collection of jobs, which produce output and information. These technologies can be heterogeneous in both the expected output produced by a worker with a given prior and the information about that worker's ability conveyed by success or failure in a job.

When a worker of any given prior is employed, a two-way trade in heterogeneous goods takes place. The worker provides labor services to the firm with perceived quality summarized by the prior. The firm provides information to the worker that is indexed by the informativeness of the assigned job. Hence, the wage varies both with the worker's heterogeneity (that is, the prior) and with the firm's heterogeneity (that is, the informativeness of the assigned job). As a result, in equilibrium, a worker may choose to work in a job that produces more information at a lower wage than he would accept at a job that produces less information. In this sense, the differing information produced by jobs leads to compensating wage differentials.

In my analysis, I assume complementarity between ability and jobs, as do MacDonald (1982) and Gibbons and Waldman (1999, 2006). I start with a simple *general ability* model in which ability is general across jobs and firms and, for simplicity, assume that jobs are equally informative. Critically, I suppose that all the jobs in the economy can be ordered according to their degree of complementarity with ability. Then, equilibrium implies an ordering of jobs in a natural hierarchy according to a worker's comparative advantage: success leads to promotion to higher-ordered jobs in which workers are paid higher wages and high-ability workers have greater comparative advantage. Likewise, failure leads to demotion to lower-ordered jobs in which workers are paid lower wages and low-ability workers have greater comparative advantage.

In the model, the turnover of workers across firms in equilibrium depends on how jobs are distributed across firms. In the simplest such distribution, each firm has the same collection of jobs so that all firms are identical, as in Gibbons and Waldman (1999, 2006). In this case, moving up the hierarchy of jobs in the economy means moving up the job ladder within a firm. Another possible distribution of jobs across firms entails one technology consisting of a collection of the lowest-ranked jobs, another technology consisting of the next tier of jobs, and so on. Such a distribution leads to a hierarchy of firms as well as jobs: moving up the hierarchy of jobs means moving up the job ladder within a firm and then turning over to the firm on the next rung of the ladder of firms. My results are phrased so that they apply to any such distribution.

The general ability model relies on the extreme assumption that ability is perfectly general across all jobs and firms. Two strands of the literature on careers argue that such a model is inconsistent with some of the observed patterns of job mobility and wages. These patterns, however, can be qualitatively well accounted for by a *mixed ability* model in which ability is not fully general. Specifically, one strand of the literature, which started with Jovanovic (1979) and

the empirical work of Flinn (1986) and Topel (1991), argues that firm-specific capital explains two important features of the data: that the hazard of separation from a firm eventually decreases with tenure in the firm and, at the same time, that workers experience sizable returns to tenure. (See Buchinsky, Fougère, Kramarz, and Tchernis (2010) for recent estimates of returns to firm tenure.) This strand motivates me to consider a mixed ability version of the model, in which ability is *firm specific*. A newer strand, pioneered by Miller (1984) and recently advanced by Kambourov and Manovskii (2009a,b), argues that occupation-specific capital can explain two other features of the data: that the hazard of switching occupation decreases with tenure in a job or experience in the labor market and that workers enjoy substantial returns to tenure in an occupation. (See also Moscarini and Vella (2008) on the relationship between occupational mobility and experience.) This strand motivates me to consider another mixed ability version of the model, in which ability is *occupation specific*.

The firm-specific version of the model extends Jovanovic’s model to a model of careers within and across firms by allowing each firm to have a collection of jobs rather than a single job. In this case, a worker’s ability is specific to firms but general across a firm’s jobs. Thus, a worker’s state can be summarized by a vector of priors about the worker’s ability at each firm. Equilibrium job assignment, then, has a two-part form. First, at each vector of priors, an index for each firm, a modified *Gittins* index, is computed: a worker is assigned to the firm with the highest value of this index. Second, given the assigned firm based on this index, the assigned job within the firm follows a within-firm hierarchy rule. To understand how the model generates the patterns observed in the data just discussed, note that success in a given firm increases the index of that firm *only*, whereas failure lowers it. Hence, continual success leads workers to move up the hierarchy of jobs of a firm but to remain continuously employed at the same firm. Continual failure, instead, leads to demotion within a given firm until the Gittins index of the firm drops below that of the next-best firm, at which time the worker turns over to that firm.

The occupation-specific version of the model provides a theory of careers in firms and occupations. In this version, a worker has occupation-specific ability and each firm has a collection of jobs in each occupation (though not necessarily all occupations). For any given ordering of the jobs of a given occupation, the model implies a hierarchy of jobs within an occupation. In this model, equilibrium assignment follows a hierarchy rule for job assignment within an occupation and a generalized Gittins index rule for occupational choice. Depending on how general ability is across the jobs of different occupations and on how jobs of different occupations are distributed across firms, the model leads to different patterns of turnover across jobs and firms, which are consistent with the evidence on occupation-specific ability.¹

¹This version of the model bears similarities to the search model of Neal (1999). Neal shows that if workers cannot look for a new career—a career in his model corresponds to a job here—without looking for a new firm,

I then analyze the case in which jobs are differentially informative about ability, so that the speed of learning differs across jobs. Such an extension is motivated on several grounds. First, starting from Prescott and Visscher (1980) and Holmström and Tirole (1989), the argument has long been made that firms assign workers to jobs taking into account that different jobs provide different amounts of information about workers' abilities. Second, Baker, Gibbs, and Holmström (1994a,b), henceforth BGH, in their influential analysis of careers in firms, suggest that managers at the firm in their data may initially be assigned to an entry-level job that possibly contributes little to firm value but at which performance may be especially informative about a manager's ability.²

To capture these features, the paper focuses on an entry job economy in which the lowest-level job in the hierarchy is more informative than all higher-level jobs. I show that as the informativeness of the entry-level job increases, the range of priors at which a worker is assigned to that job expands. This result implies that assignment is no longer dictated by the static comparative advantage of workers to jobs. Indeed, there exists a range of priors at which workers are assigned to jobs at which they have static comparative disadvantages but are *dynamically* compensated for this assignment with a wage differential when promoted out of such a job. Specifically, this compensating differential reflects the value of the loss in information about ability that workers incur when they move from a job of *higher* informativeness to a job of *lower* informativeness and leads to discontinuous increases in wages upon a job change. This logic provides a novel argument for the observation that individuals commonly experience sizeable wage increases at promotion. In particular, the model can explain why wage increases following promotion may rise with hierarchical level and experience, as is the case in the BGH data, even when much information has already been acquired about a worker. Commonly, learning models in which jobs are equally informative cannot well explain these feature of the data. (See the discussion in Gibbons and Waldman (2006) of the difficulty of learning models in accounting for these facts.)

To illustrate these predictions, I present a simple two-period example in which the informativeness of a job decreases with the level of the job in the firm hierarchy. In equilibrium, successful workers are promoted by one level at most and experience wage increases, which are discontinuous (in the prior) for workers with high enough priors at a level promoted to a different level. Unsuccessful workers, instead, experience wage decreases but are never demoted. These features of the model are consistent with commonly observed characteristics of careers in firms and, in particular, with those in the BGH data.

then a two-stage search strategy is optimal, whereby workers look first for their best career matches and then for their best firm matches. His model, however, abstracts from learning (and human capital acquisition) and focuses on job mobility only.

²In Pastorino (2013) I provide formal evidence supporting this conjecture in the BGH data. Specifically, I estimate a labor market model with job assignment, learning, and human capital acquisition using the BGH data and find that the entry-level job in the firm's hierarchy has greater informativeness than higher-level jobs.

Although the proposed model is rich enough to be promising in accounting for many observed features of careers, one dimension it may not capture is that individual wages *on average* increase with experience in the labor market, a feature that a competitive learning model with complementarity between ability and jobs can explain only upon a job or firm switch. (See Gibbons and Waldman (1999) for a discussion of this point.) I show that human capital can be straightforwardly incorporated into the model to account for this missing feature. Moreover, equilibrium in this augmented model displays the same qualitative patterns of job assignment as before.

Related Literature. The model is developed around the idea that important complementarities exist between workers and jobs. This specification of the production technology of a firm parallels those found in the literature (see Waldman (1984), Bernhardt (1995), Gibbons and Waldman (1999, 2006), and the references therein cited).³ Following Jovanovic (1979), Prescott and Visscher (1980), MacDonald (1982), Miller (1984), and Gibbons and Waldman (1999, 2006), I model informational human capital by allowing firms and workers to learn (symmetrically) about a fixed set of productive skills of a worker, referred to as *ability*. The assumption of symmetric learning based on a worker's output is common in the literature on learning in the labor market.

I allow for ability to be firm specific to a varying degree, in order to nest the cases of purely firm-specific ability as in Jovanovic (1979), Miller (1984), and Flinn (1986); purely general ability, across jobs and firms, as in MacDonald (1982) and Gibbons and Waldman (1999, 2006); and ability that is imperfectly transferable across jobs or firms. When ability is general across a firm's jobs, the production structure of my model shares the feature of the one in Rosen (1982), Waldman (1984), and Gibbons and Waldman (1999, 2006) that tasks or jobs can be ordered in terms of the extra value that a worker of greater ability produces. When ability is imperfectly transferable across jobs, the model provides a theory of sorting within and across occupations, whose patterns depend on the distribution of skills and jobs across firms. Jovanovic and Nyarko (1997) also examine patterns of occupational mobility between occupations but focus on contrasting the implications of a pure learning model to those of a learning-by-doing model in which what is learned about production in an occupation is also valuable for production in another occupation. Along this dimension, my work is related to that of Groes, Kircher, and Manovskii (2010), who analyze patterns of occupational mobility among low and high earners in an occupation. Basing their analysis on a theory of learning about absolute advantage, they provide an explanation for the U-shaped pattern of turnover across

³In these papers a worker's productivity is assumed to be independent of other workers' assignments within the firm. For papers that relax this assumption, see Kremer (1993), Kremer and Maskin (1996), and, more recently, Ferrall (1997) and Ferrall, Salvanes, and Sørensen (2009), who all analyze the impact of complementarity in production among workers on the assignment of workers to tasks or teams. The models in these papers, however, are silent about the dynamics of job assignments and wages, which is the focus of the analysis here.

occupations. This U-shaped pattern is generated by the fact that the probability of leaving an occupation is highest at the lowest and highest percentiles of the distribution of wages in an occupation.

My paper is also related to the learning models of MacDonald (1982), Felli and Harris (2004), Eeckhout and Xi (2009), and Anderson and Smith (2010). MacDonald (1982) analyzes equilibrium assignment in a competitive economy with task complementarities in which workers' abilities are unobserved, but assumes that the jobs of all firms are equally informative about ability and that ability is purely general. Felli and Harris (2004) consider an oligopoly setting in which the jobs of a firm differ in the amount of information they provide. They focus on the case in which ability is firm specific and do not characterize equilibrium assignment. Eeckhout and Xi (2009) and Anderson and Smith (2010) examine matching in the labor market with two-sided heterogeneity and learning. Whereas Anderson and Smith (2010) focus on limiting conditions under which assortative matching may or may not arise, Eeckhout and Xi (2009) characterize equilibrium assignment when firms consists of two jobs and ability is purely general across jobs and firms.

1 The Economy

Here, I set up the model and then discuss its information structure.

1.1 Setup

Consider a market in which N firms compete for workers in each period of an infinite (or finite) horizon, with dates $t \geq 1$. All firms and workers share a common discount factor, $\delta \in [0, 1)$. In the economy, only one good is produced and consumed, the price of which is normalized to one. A worker's talent or productive ability at each firm f is described by the skill type $s \in \{a, b\}$, which is assumed to be unobserved by both firms and the worker. I refer to a worker of type a as a *high-ability* worker and a worker of type b as a *low-ability* worker. Each worker inelastically supplies one unit of labor each period. At the initial date, $t = 1$, the information about any worker is summarized by the prior probability p_1 that this worker is of high ability. I assume there is a continuum of workers with a distribution of initial priors $G(p_1)$ and associated density $g(p_1)$.

Each firm is endowed with a production set that consists of K_f tasks or *jobs*. A worker's output in a period is stochastic, and its distribution depends on the worker's ability and the job that the worker performs. The output produced in period t , when the worker is employed at job k of firm f , $1 \leq k \leq K_f$, is denoted by $y_{fk} \in \{y_{fHk}, y_{fLk}\}$, with $y_{fHk} > y_{fLk} > 0$. For

a worker of ability a , the probability of high output at job k of firm f is α_{fk} , whereas for a worker of ability b , the probability of high output at job k of firm f is β_{fk} , with

$$0 < \beta_{fk} < \alpha_{fk} < 1. \quad (1)$$

Thus, a job is a four-tuple $\{\alpha_{fk}, \beta_{fk}, y_{fHk}, y_{fLk}\}$. A worker's output, or performance, is symmetrically observed by all firms and workers. Assumption (1) implies that learning about ability is gradual ($\alpha_{fk}, \beta_{fk} \in (0, 1)$) and that the output distribution at each job satisfies the monotone likelihood-ratio property ($\alpha_{fk} > \beta_{fk}$); that is, observing high output, y_{fHk} , increases the probability that the firms and the worker assign to the worker's ability being high.

At the beginning of period $t \geq 1$, the one-period expected output when the worker is assigned to job k of firm f , conditional on ability being high and low, respectively, is

$$\bar{y}_f(a, k) = \alpha_{fk}y_{fHk} + (1 - \alpha_{fk})y_{fLk} \text{ and } \bar{y}_f(b, k) = \beta_{fk}y_{fHk} + (1 - \beta_{fk})y_{fLk}. \quad (2)$$

The one-period expected output $y_f(p_t, k)$ at job k of firm f in $t \geq 1$ given p_t is

$$y_f(p_t, k) = p_t\bar{y}_f(a, k) + (1 - p_t)\bar{y}_f(b, k). \quad (3)$$

Notice that $y_{fHk} > y_{fLk}$ and (1) imply that the expected output $y_f(p_t, k)$ at any job k of firm f is *monotone* in the prior p_t . I assume throughout that

$$\bar{y}_f(a, k + 1) \geq \bar{y}_f(a, k) \text{ and } \bar{y}_f(b, k) \geq \bar{y}_f(b, k + 1), \quad k = 1, \dots, K_f - 1. \quad (4)$$

To interpret this assumption, suppose that managerial talent at each firm f corresponds to the ability to implement successfully risky projects at any job k , where the probability of success at each project is α_{fk} if the worker's ability is high and β_{fk} if the worker's ability is low. This assumption implies that if projects are ranked by their complexity, from the simplest (job 1) to the most involved (job K_f), then more complex projects are those at which complementarity with managerial talent is more pronounced, since (4) implies that $\bar{y}_f(a, k + 1) - \bar{y}_f(a, k) \geq \bar{y}_f(b, k + 1) - \bar{y}_f(b, k)$. In particular, whereas high-ability workers are more productive at higher-level jobs, low-ability workers are better suited to lower-level ones. Notice that (1) and (4) also imply *single crossing*, namely, that for any two jobs k and k' of firm f ,

$$y_f(p, k) \text{ and } y_f(p, k') \text{ cross once in } p \text{ for } p \in [0, 1]. \quad (5)$$

See Figure 1 for static rewards that satisfy these assumptions.

Starting from the initial prior p_1 for a worker, the prior is updated in the obvious recursive

way using Bayes' rule. In particular, if p_t is the prior that the worker is of high ability at the beginning of period t , and the worker is assigned to job k of firm f , then if the realized output is y_{fHk} , the updated prior p_{t+1} at the beginning of period $t + 1$ is

$$P_H(p_t, \theta_{fk}) = \frac{\alpha_{fk} p_t}{\alpha_{fk} p_t + \beta_{fk} (1 - p_t)} \quad (6)$$

with $\theta_{fk} = (\alpha_{fk}, \beta_{fk})$. If the realized output is y_{fLk} , the updated prior p_{t+1} at the beginning of period $t + 1$ is

$$P_L(p_t, \theta_{fk}) = \frac{(1 - \alpha_{fk}) p_t}{(1 - \alpha_{fk}) p_t + (1 - \beta_{fk}) (1 - p_t)}. \quad (7)$$

I will say that job k is *more informative* than job k' if the posterior reached after output is realized at job k is a mean-preserving spread of the posterior reached after output is realized at job k' in that

$$P_H(p_t, \theta_{fk}) \geq P_{Hk'}(p_t, \theta_{fk'}) \text{ and } P_{Lk}(p_t, \theta_{fk}) \leq P_{Lk'}(p_t, \theta_{fk'}), \quad (8)$$

as shown in Figure 2. For convenience, I define the updating rule after $n \geq 2$ consecutive successes and failures, respectively, as $P_H^n(p_t, \theta_{fk})$ and $P_L^n(p_t, \theta_{fk})$.

1.2 Information Structure

Here, I model the distribution of output at each job as discrete and possibly asymmetric (in that β_{fk} need not equal $1 - \alpha_{fk}$). Most of the literature, including Jovanovic (1979) and Gibbons and Waldman (1999, 2006), model the signals as continuous (normally distributed) and symmetric. My setup has several advantages. First, the performance data from firm personnel records are discrete in that they come in a coarse scale. In most cases, this scale simply has two values, such as satisfactory or not; in others, it takes only a few other values. Moreover, it is well known that scores are compressed, so that even if the scale allows, say, for three or more scores, nearly all of the scores are one of two values. (See the discussion in Pastorino (2013).) Second, my discrete Bernoulli type of information structure allows for flexible time variation of the variance of posterior beliefs, and thus wages and job assignment. In contrast, the normal type of information structure, commonly used, implies that the variance of posterior beliefs declines deterministically in the number of signals observed. Hence, this structure in a competitive economy would imply that the variance of equilibrium wages also declines monotonically with tenure. This pattern, however, is not universal. For example, it is not present in the BGH data. As such, in practice, a discrete distribution may be empirically more relevant.

Finally, the normal type of information structure imposes symmetry in the distribution of

output realizations across types. This symmetry constrains the pattern of belief updating and, in turn, the patterns of promotions and demotions. In contrast, the discrete signal structure assumed here naturally accommodates asymmetric output distributions between workers of high and low ability. These asymmetric distributions can more easily account for the common observation in the data on careers in firms that there are virtually no demotions and promotions are gradual. The asymmetric model can generate these observations under two conditions. First, a high-ability and a low-ability worker have a small chance of low output, so demotions are rare. Second, a high-ability worker's chance of high output is only slightly higher than that of a low-ability worker, so the information produced by a success is small. If so, then promotions are gradual. Below, I discuss an example that makes this point.

2 A Competitive Equilibrium

Here, I set up a competitive equilibrium for this economy, first by defining notation for the goods being exchanged. Note that when a worker supplies labor to a firm in a given job, there is two-sided heterogeneity, on the part of the worker and on the part of the job. Worker heterogeneity is due to workers being of different unobserved skill types and is summarized by the prior p_t , which may be thought of as the worker's *information capital* at date t . Thus, a worker with prior p_t supplies a labor input that is different from that of a worker with prior p'_t . Job heterogeneity is due to the fact that each job k of any firm f supplies a possibly different increment to this information capital. Specifically, if a worker supplies labor in a job indexed by θ_{fk} , then the worker's information capital is (stochastically) augmented according to (6) and (7). I refer to θ_{fk} as the *informativeness* of job k at firm f . The updating rules in (6) and (7) are essentially production functions for incremental information that a firm supplies to a worker by hiring the worker in a given job. Because of this two-sided heterogeneity, a worker's wage at t will depend on the observable characteristics of both the worker, p_t , and of the job the worker is assigned to, θ_{fk} . Hence, at any time t the wage function is $w_t(p_t, \theta_{fk})$.

Now consider the problem of the representative firm f that takes this wage function as given. This firm maximizes the present discounted value of profits by solving

$$\Pi_{ft} = \max_{\{l_{ft}(p_t, k)\}_k} \left\{ (1 - \delta) \sum_{k=1}^{K_f} \int [y_f(p_t, k) - w_t(p_t, \theta_{fk})] l_{ft}(p_t, k) dp_t + \delta \Pi_{ft+1} \right\}, \quad (9)$$

where $l_{ft}(p_t, k)$ is the measure of workers with prior p_t that firm f hires in job $k \in K_f$ at time t . Here, as in the worker's problem below, I normalize payoffs by $1 - \delta$ so as to express them as per-period averages. Since the firm rents labor services each period, there are no state variables for the firm, and the dynamic problem (9) reduces to a sequence of static problems. Each static

problem reduces to a job-by-job maximization problem for workers of each level of information capital p_t , given by

$$\max_{\{l_{ft}(p_t, k)\}_k} [y_f(p_t, k) - w_t(p_t, \theta_{fk})] l_{ft}(p_t, k). \quad (10)$$

Notice that since the firm is hiring a measure of workers and all of the uncertainty about a worker's production is idiosyncratic, the law of large numbers implies that there is no uncertainty at the firm level.

Next consider the problem of a worker with initial prior p_1 . Each such worker has endowments of one unit of labor services in each period and owns a per capita share of aggregate profits. Each worker has preferences over consumption c_t given by $(1 - \delta)E_1 \sum_{t=1}^{\infty} \delta^{t-1} c_t$. The income of each worker is the sum of wage income and profit income π_t , which is the sum of the period t profits of the N firms. I assume there are no asset markets so that in each period a worker sets consumption equal to income.⁴ In a worker's problem, the aggregate profits π_t are simply an additive constant to consumption that has no effect on decisions. Hence, I do not explicitly include them in the worker's problem. Since they will be zero in equilibrium, they also have no effect on the value of a worker's utility.

Now letting d_{fkt} be an indicator function for whether the worker chooses to work at job k of firm f at t ($d_{fkt} = 1$) or not ($d_{fkt} = 0$), with $\sum_{fk} d_{fkt} = 1$, I can write this problem as

$$V_t^w(p_t) = \max_{\{d_{fk}\}_f} \sum_{f=1}^N \sum_{k=1}^{K_f} d_{fk} [(1 - \delta)w_t(p_t, \theta_{fk}) + \delta EV_{t+1}^w(p_{t+1}|p_t, \theta_{fk})], \quad (11)$$

where $EV_{t+1}^w(p_{t+1}|p_t, \theta_{fk}) = r(p_t, \theta_{fk})V_{t+1}^w(P_H(p_t, \theta_{fk})) + [1 - r(p_t, \theta_{fk})]V_{t+1}^w(P_L(p_t, \theta_{fk}))$ is the continuation value, given the current state p_t and the current assigned job k of firm f , and $r(p_t, \theta_{fk}) = \alpha_{fk}p_t + \beta_{fk}(1 - p_t)$ is the probability of success when the worker is in such a job. Let $d_t = \{d_{fkt}(p_t)\}$ denote the decision rule of a worker at time t . Starting from any initial distribution of priors $G_1(p_1)$, these decision rules imply a transition law from $G_t(p_t)$ to $G_{t+1}(p_{t+1})$ of the form

$$G_{t+1}(\cdot) = \Gamma(G_t(\cdot), d_t). \quad (12)$$

Then, the market clearing condition for labor is that at each $p \in [0, 1]$ and time period,

$$\int_0^p d_{fkt}(p_t) dG_t(p_t) = \int_0^p l_{fkt}(p_t) dp_t \quad (13)$$

at the job of each firm. For intuition, imagine that G_t is a discrete distribution so that $g_t(p_t)$

⁴I show in the Appendix that this economy is equivalent to one in which workers are risk averse with utility function $u(c_t)$, there are complete markets against idiosyncratic risk, and the economy is stationary as workers stochastically die and new ones enter.

is the fraction of workers with prior p_t . Then, (13) requires that at every prior p_t , the fraction of workers who choose to work in job k of firm f in period t , namely, $d_{fkt}(p_t)g_t(p_t)$, equals the fraction of the labor force demanded at that job, namely, $l_{fkt}(p_t)$. Notice that the time subscript on all variables is due to the dependence on the distribution of priors G_t , which evolves with time.

A *competitive equilibrium* is a sequence of wage functions $\{w_t(p_t, \theta_{fk})\}$, worker job choices $\{d_{fkt}(p_t)\}$, firm labor choices $\{l_{fkt}(p_t)\}$, and distributions of priors $\{G_t(p_t)\}$ such that: (i) given the wage functions, each worker's job choices solve the worker problem (11), (ii) given the wage functions, each firm's labor choices solve the firm problem (10), (iii) the worker decision rules are consistent with the transition law (12) for the distribution of priors, (iv) each firm makes zero profits at each date, and (v) the labor market clears.

Turning to the characterization of equilibrium, first consider wages. The zero profit condition immediately implies that $w_t(p_t, \theta_{fk}) \geq y_f(p_t, k)$ for all jobs, priors, and dates, and

$$w_t(p_t, \theta_{fk}) = y_f(p_t, k) \quad (14)$$

whenever $l_{fkt}(p_t) > 0$. Notice first that (14) implies that wages do not vary with time, hence I drop the time subscript on wages from now on. Next, notice that there are many ways of setting wages for unchosen jobs (those with $l_{fkt}(p_t) = 0$) that are consistent with the same allocations. Here, I do so as follows. Recall that an arbitrary $\theta_{fk} = (\alpha_{fk}, \beta_{fk})$ indexes the informativeness of job k of firm f . Let $A(\theta)$ denote the set of output levels (y_{fHk}, y_{fLk}) of jobs $(\alpha_{fk}, \beta_{fk}, y_{fHk}, y_{fLk})$ in all the N firms such that $(\alpha_{fk}, \beta_{fk}, y_{fHk}, y_{fLk}) = (\theta, y_{fHk}, y_{fLk})$ for the given θ , that is,

$$A(\theta) = \{(y_{fHk}, y_{fLk}) \mid (\alpha_{fk}, \beta_{fk}, y_{fHk}, y_{fLk}) = (\theta, y_{fHk}, y_{fLk})\},$$

and let

$$w(p_t, \theta) = \max_{(y_{fHk}, y_{fLk}) \in A(\theta)} y_f(p_t, k). \quad (15)$$

I refer to $w(p_t, \theta)$ as the equilibrium *wage function*. To help with the intuition for (15), suppose only two jobs in the economy have a given level of θ , say, job k of firm f and job k' of firm f' with $\theta_{fk} = \theta_{f'k'} = \theta$. Now suppose that for $p < \bar{p}$, $y_f(p, k) > y_{f'}(p, k')$ and for $p \geq \bar{p}$, $y_f(p, k) \leq y_{f'}(p, k')$. Then, the wage equals

$$w(p_t, \theta) = \left\{ \begin{array}{l} y_f(p, k) \text{ for } p < \bar{p} \\ y_{f'}(p, k') \text{ for } p \geq \bar{p} \end{array} \right\}.$$

At these wages, for $p < \bar{p}$, firm f will make zero profits at job k and firm f' will make negative

profits at job k' , whereas for $p \geq \bar{p}$, firm f will make negative profits at job k and firm f' will make zero profits at job k' . The *planning problem* for a given worker in this economy can now be stated:

$$W(p_t) = \max_{\{d_{fk}\}} \sum_{f=1}^N \sum_{k=1}^{K_f} d_{fk} [(1 - \delta)y_f(p_t, k) + \delta EW(p_{t+1}|p_t, \theta_{fk})], \quad (16)$$

where at $t = 1$, the initial prior p_1 is given. Notice that the planning problem for a given worker is a stationary recursive problem. In this economy with linear utility, the intertemporal marginal rate of substitution in consumption is constant at δ and, hence, (16) applies even when the prior distribution G_t is time varying.⁵ The following result can be established.

Proposition 1. *The competitive equilibrium is efficient in that for each worker with initial prior p_1 , market-wide job assignments solve (16) and each worker is paid the expected output of the resulting job as in (15).*

Proof: To prove this proposition, denote the set of informativeness levels θ of all jobs in the economy by $\Theta = \{\theta | \exists f, k: \theta_{fk} = \theta\}$. Then using (15), the worker's problem can be written as

$$V^w(p_t) = \max_{\theta \in \Theta} [(1 - \delta)w_t(p_t, \theta) + \delta EV^w(p_{t+1}|p_t, \theta)], \quad (17)$$

which using (15) can be written as

$$V^w(p_t) = \max_{\theta \in \Theta} \left[(1 - \delta) \max_{(y_{fHk}, y_{fLk}) \in A(\theta)} y_f(p_t, k) + \delta EV^w(p_{t+1}|p_t, \theta) \right] \quad (18)$$

or, since the second term in (18) does not depend on (y_{fHk}, y_{fLk}) , equivalently

$$V^w(p_t) = \max_{\theta \in \Theta} \left\{ \max_{(y_{fHk}, y_{fLk}) \in A(\theta)} [(1 - \delta)y_f(p_t, k) + \delta EV^w(p_{t+1}|p_t, \theta)] \right\}, \quad (19)$$

which clearly equals (16). Notice that (16) simply adds to (19) the option of choosing dominated jobs for a given θ . \square

Notice that when θ_{fk} varies across firms and jobs, the planning problem (16) is a multi-armed bandit problem with dependent arms in which the planner (and the worker in equilibrium) trades off the value of current output in a job against the informativeness of the job. In general, the solution to this problem does not imply that the job with the highest current expected output is chosen.⁶ It will prove convenient notationally to write the problem in (16)

⁵When the economy is interpreted as an overlapping generations economy with risk-averse consumers, (16) applies only to the stationary distribution; otherwise, the discount factor δ will be replaced by the gross interest rate, $1/R_t$, and hence will vary with time.

⁶The bandit problem describes the sequential sampling problem of a decision maker uncertain about the

as

$$W(p_t) = \max_{f,k \in K_f} [(1 - \delta)y_f(p_t, k) + \delta EW(p_{t+1}|p_t, \theta_{fk})]. \quad (20)$$

3 General Ability

Here, I assume that a worker's ability is general across jobs and firms. In this economy, each firm f has K_f jobs, so there is a total of $\bar{K} = \sum_{f=1}^N K_f$ jobs. Suppose that these jobs can be ordered by a single index l , say, $l = l(f, k)$, such that these jobs satisfy the analog of (4) for the entire set of \bar{K} jobs in that

$$\bar{y}(a, l + 1) \geq \bar{y}(a, l) \text{ and } \bar{y}(b, l) \geq \bar{y}(b, l + 1), \quad l = 1, \dots, \bar{K} - 1, \quad (21)$$

where it is understood that $\bar{y}(p, l(f, k)) = \bar{y}_f(p, k)$. Notice that (21) is a strengthening of (4) in that even if (4) holds for the jobs of each firm, it does not imply that (21) holds for the jobs of all firms. For simplicity, I also assume that informativeness is the same across all firms and jobs in that

$$\alpha_{fk} = \alpha \text{ and } \beta_{fk} = \beta \text{ for all } f \text{ and } k, \quad (22)$$

so that $\theta_{fk} = \theta$ is identical across all firms and jobs. Therefore, I suppress θ_{fk} in the following whenever unambiguous. I will later relax these assumptions.

My characterization applies to the set of all jobs \bar{K} for any such ordering $l(f, k)$. Given this ordering, the implications of the model for turnover between jobs in a firm and between jobs across firms depends only on how these jobs are distributed across firms.

For ease of exposition only, I first suppose that all firms are identical so that $l(f, k) = l(f', k)$ and the ordering $l(f, k)$ reduces to an order on the $K (= \bar{K}/N)$ jobs of each such firm, indexed by $k = 1, \dots, K$. For this reason, the firm subscript below is suppressed. I then discuss alternative orderings and their implications for turnover across firms. Note that regardless of the distribution of jobs across firms, in this economy job assignment is purely driven by output considerations, since, unlike in the general model analyzed below, the assigned job does not affect the amount of information generated about a worker's ability.

I start with a characterization of the job assignment rule and the associated wages, and then examine the implications for promotions and demotions.

distribution of payoffs of the available alternatives. Here, each arm corresponds to a job. Since the worker's ability is (perfectly) correlated across jobs, the returns from the arms, here the jobs of all firms, are dependent.

3.1 Characterization

I start by showing that restriction (22) implies that the equilibrium job assignment rule and paid wages are both static, and that workers with higher priors are assigned to higher-indexed jobs. In particular, the equilibrium wage function reduces to

$$w(p_t) = \max_k y(p_t, k), \quad (23)$$

and a worker with prior p_t chooses the job k and firm f that lead to (23). Define the job assignment rule $k^*(p_t)$ as

$$k^*(p_t) = \arg \max_{k \in K} y(p_t, k). \quad (24)$$

Proposition 2. *Under (4), when firms are identical in their technologies and jobs are equally informative as in (22), the equilibrium job assignment rule is static in that a worker is assigned to the job with the highest current expected output and a worker's wage equals the worker's expected output at the employing firm. Moreover, $k^*(p)$ is increasing in p .*

Proof: By Proposition 1, equilibrium job assignment solves problem (20). Since firms are identical, the subscript f can be dropped, and since all jobs are equally informative, (20) reduces to

$$W(p_t) = (1 - \delta) \max_{k \in K} [y(p_t, k)] + \delta EW(p_{t+1}|p_t), \quad (25)$$

which implies that the assignment rule solves $\max_{k \in K} y(p_t, k)$. By (14), it also follows that the equilibrium wage is given by (23). To prove that the optimal policy is increasing in p , note that by monotonicity of expected output in p and by single crossing, if a higher-indexed job k' is preferred to a lower-indexed job k at some prior p , then k' must also be preferred to k at any higher prior $p' \geq p$. \square

Note that in this equilibrium, workers are indifferent about which firm they choose. If we think of there being an infinitesimal cost of switching firms either on the worker side or on the firm side, then there is no turnover in equilibrium.

This benchmark model implies a number of features of job and wage dynamics in line with the patterns observed in firms. These features can be illustrated through a complete characterization of the path implied by the optimal job assignment rule phrased in recursive form. To do so, let $k_H(p)$ and $k_L(p)$ be shorthand notation for $k^*(P_{Hk^*(p)}(p))$ and $k^*(P_{Lk^*(p)}(p))$, respectively, the job assigned after a success at the optimal job at prior p and the job assigned after a failure at the optimal job at prior p . The next corollary provides this characterization.

Corollary 1. *Under (4), the optimal job assignment rule is: (i) in any period t , if the prior is p assign $k^*(p)$, (ii) upon a success, in period $t + 1$ update the prior to $P_{Hk^*(p)}(p)$ and assign job*

$k_H(p) \geq k^*(p)$, (iii) upon a failure, in period $t + 1$ update the prior to $P_{Lk^*(p)}(p)$ and assign job $k_L(p) \leq k^*(p)$.

A convenient way to visualize the assignment rule implied by this proposition is as a partition of the set of priors. To do so, let ρ_k , $k = 2, \dots, K$, denote the cutoff prior at which expected output is equal at jobs $k - 1$ and k , defined by $y(p_t, k - 1) = y(p_t, k)$ and given by

$$\rho_k = \frac{y(b, k - 1) - y(b, k)}{y(b, k - 1) - y(b, k) + y(a, k) - y(a, k - 1)}. \quad (26)$$

Notice that assumption (4) allows for the possibility that some job is dominated by the others in that it is not optimal at any prior. (For example, if, say, job 3 is dominated, then as the prior increases, the worker might move from job 2 to job 4, skipping over job 3.) If dominated jobs are ruled out, by assuming that for each job k , there exists a prior $p(k)$ such that

$$y(p(k), k) > y(p(k), k') \text{ for all } k' \neq k, \quad (27)$$

then the optimal assignment rule can be represented by a partition of the set of priors into disjoint intervals such that at prior $p_t \in [\rho_k, \rho_{k+1})$, a worker is assigned to job k , where

$$0 < \rho_k < \rho_{k+1} < 1. \quad (28)$$

Here, I have adopted the convention that if two jobs are tied in expected output, then the higher-indexed job is selected. I refer to this partition as the *interval characterization* of the optimal assignment rule. The wages of a worker with prior p_t are

$$w^*(p_t) = y(p_t, k^*(p_t)). \quad (29)$$

The following corollary summarizes the patterns generated by the equilibrium assignment and wage rules. The proof is relegated to the Appendix.

Corollary 2. *In the benchmark model: (i) wages increase both upon promotion and after successful performance, (ii) wages decrease both upon demotion and unsuccessful performance, (iii) workers assigned to higher-level jobs are paid higher wages, (iv) conditional on current wages, wage increases are positively serially dependent, (v) conditional on current wages, a worker promoted from job k to job k' is more likely than an unpromoted worker to receive a future promotion to job k'' , and (vi) wage increases at promotion can be small relative to the range of possible wages at a level.*

So far I have focused on the paths of wages and job assignments for a single worker. When there is a nondegenerate distribution of workers across jobs, this model also has implications

for how certain cross-sectional moments of a cohort of workers vary with time. This is first illustrated with a simple result on the limiting distribution of workers across jobs. Next I present an example of how heterogeneous priors allow the model to account for various patterns of promotions observed in the data.

For the first result, consider an initial prior distribution $G_1(p)$ with *interior* support, that is, a support $[\underline{p}, \bar{p}]$ with $0 < \underline{p} < \bar{p} < 1$. This result immediately follows by standard properties of martingales.

Corollary 3. *Suppose (4) holds and consider an initial prior distribution with interior support. Then, in the limit under the optimal assignment rule, a measure $\int_{\underline{p}}^{\bar{p}} p dG_1(p)$ of workers are assigned to job K and a measure $\int_{\underline{p}}^{\bar{p}} (1 - p) dG_1(p)$ of workers are assigned to job 1.*

The intuition for this corollary is that at any point in time, the job assignment process provides a strictly positive amount of information at any interior prior. Hence, by repeatedly sampling observations on the performance of each worker, eventually the true ability of all workers is learned.

3.2 Promotions and Demotions

I now turn to a detailed discussion of the implications of the model for promotions and demotions. I begin with an example motivated by a striking feature of the BGH data, namely, that the hazard rate of promotion is nonmonotone in tenure. In particular, the probability of promotion from one job level to the next first increases, then decreases with tenure. I show here that the benchmark model can easily generate this pattern for a suitable nondegenerate distribution of initial priors $G(p_1)$.

I then refine the interval-based rule for job assignment, showing that equilibrium implies that the interval of priors associated with any job k , namely, $[\rho_k, \rho_{k+1})$, can be further partitioned into *rungs of promotion and demotion*. Specifically, a success at prior p in this interval leads to a new prior $P_H(p)$. In general, this new prior could remain in that interval, so that the worker stays in job k , could land in the interval $[\rho_{k+1}, \rho_{k+2})$, so that the worker is promoted by one level to job $k + 1$, or could land in some higher interval $[\rho_{k+r}, \rho_{k+r+1})$, so that the worker is promoted by r levels. Likewise, a failure at a prior in this interval could lead the worker to stay in job k , be demoted by one level, or be demoted by more than one level. I begin this refinement by characterizing through a partition of $[\rho_k, \rho_{k+1})$ the number of successes and failures required for a promotion and a demotion by exactly one level. I then consider multi-level promotions and demotions.

3.2.1 Heterogeneous Priors and Promotion Patterns

For simplicity only, assume that the output distribution is symmetric across types of workers in that type a workers succeed with probability α and type b workers succeed with probability $\beta = 1 - \alpha$ at any job. This symmetry ensures that one failure and one success cancel each other in that $P_H(P_L(p)) = P_L(P_H(p)) = p$. Next, in a slight abuse of notation, let the initial distribution of priors be discrete and correspond to just two groups of workers, denoted by A and B , with mass $g(p_A)$ on prior p_A and mass $g(p_B)$ on prior p_B , respectively, where $p_B > p_A$. Suppose the initial priors are such that both groups start at job 1 in period 1 but that group B workers need one success to be promoted, whereas group A workers need two successes, that is,

$$P_H(p_A) < \rho_2 < P_H(p_B) \quad (30)$$

and $P_H(P_H(p_A)) > \rho_2$, where ρ_2 is the cutoff prior for promotion to job 2. This example is illustrated by Figure 3.

After one year, only the group B workers that were successful get promoted. The successful group A workers have updated prior $P_H(p_A)$, which by (30) falls short of the promotion cutoff. Hence, after one year the fraction

$$r(p_B)g(p_B) \quad (31)$$

of all workers get promoted, where $r(p) = \alpha p + \beta(1 - p)$ denotes the fraction of successes for a group with prior p .

After two years, only the group A workers who received two successes get promoted, so the fraction of promoted workers is

$$r(P_H(p_A))r(p_A)g(p_A). \quad (32)$$

The group B workers that failed in the first year and then succeeded in the second have priors $P_H(P_L(p_B))$, which by symmetry equals the initial prior p_B , which is below the cutoff for promotion. Likewise, the group B workers that succeeded in the first year and then failed in the second have priors $P_L(P_H(p_B))$, which also equal p_B . Adding together these two sets of group B workers gives that after two years there is a measure of

$$\{r(p_B)[1 - r(P_H(p_B))] + [1 - r(p_B)]r(P_L(p_B))\}g(p_B) \quad (33)$$

workers at prior p_B .

After three years, the group B workers of measure (33) with priors p_B who receive a success have priors $P_H(p_B)$. By (30) these workers are promoted, and using (33) these workers are fraction

$$r(p_B)\{r(p_B)[1 - r(P_H(p_B))] + [1 - r(p_B)]r(P_L(p_B))\}g(p_B) \quad (34)$$

of the workers. A moment's reflection makes clear that no workers from group A can be promoted in period 3. The reason is that any group A worker with three successes has already been promoted in period 2, and any group A worker who experienced at least one failure has a prior no greater than $P_H(P_H(P_L(p_A))) = P_H(p_A)$, which by (30) falls short of the cutoff level of promotion ρ_2 . Clearly, if $g(p_A)$ is sufficiently larger than $g(p_B) = 1 - g(p_A)$, then the promotion rate initially increases and then decreases, in that (31) is smaller than (32), which is larger than (34).

As this simple example should make clear, with multiple groups of workers of varying proportions and with different priors, the model can produce very flexible tenure profiles of the hazard rate of promotion out of a job.

3.2.2 Steps to Promotion and Demotion by One Level

In the BGH data, nearly all of the promotions are by one level, and (the relatively few) demotions are also by one level. Motivated by this observation, I begin with the simple case in which

$$P_H(\rho_{k+1}) < \rho_{k+2} \text{ and } P_L(\rho_k) \geq \rho_{k-1} \quad (35)$$

so that all promotions and demotions are by one level. The more general case is discussed later. Under (35), I define a *promotion rung* as an interval of priors at which the same number of consecutive successes leads to promotion by one level and a *demotion rung* as an interval of priors at which the same number of consecutive failures leads to a demotion by one level.

Now promotion rungs can be constructed. Given a cutoff prior ρ_{k+1} for promotion to job $k + 1$, define the priors $\bar{q}_{k1}, \bar{q}_{k2}, \dots$, by $P_H(\bar{q}_{k1}) = \rho_{k+1}$, $P_H^2(\bar{q}_{k2}) = \rho_{k+1}$, and so on, where $P_H^2(p)$ denotes $P_H(P_H(p))$ and similar notation is used for $P_H^n(p)$. Let \bar{n}_k be the smallest integer such that $P_H^{\bar{n}_k}(\rho_k) \geq \rho_{k+1}$. Then job k has \bar{n}_k promotion rungs given by

$$[\rho_k, \bar{q}_{k, \bar{n}_k - 1}), \dots, [\bar{q}_{k2}, \bar{q}_{k1}), [\bar{q}_{k1}, \rho_{k+1}) \quad (36)$$

with the interpretation that a worker in promotion rung $n \in \{1, \dots, \bar{n}_k\}$ of job k , namely, with priors in $[\bar{q}_{kn}, \bar{q}_{k(n-1)})$, will be promoted exactly after n successes.

Demotion rungs are constructed analogously. Given a cutoff prior ρ_k for demotion from job k to job $k - 1$, define the priors $\underline{q}_{k1}, \underline{q}_{k2}, \dots$, by $P_L(\underline{q}_{k1}) = \rho_k$, $P_L^2(\underline{q}_{k2}) = \rho_k$, and \underline{n}_k be the smallest integer such that $P_L^{\underline{n}_k}(\rho_{k+1}) < \rho_k$. (The equalities $P_L^n(\underline{q}_{kn}) = \rho_k$ are intended to hold for priors just below the cutoff ρ_k .) Then job k has \underline{n}_k demotion rungs given by

$$[\rho_k, \underline{q}_{k1}), [\underline{q}_{k1}, \underline{q}_{k2}), \dots, [\underline{q}_{k, \underline{n}_k}, \rho_{k+1}) \quad (37)$$

with the interpretation that a worker in demotion rung $n \in \{1, \dots, \underline{n}_k\}$ of job k , namely, with a prior in $[\underline{q}_{kn-1}, \underline{q}_{kn})$, will be demoted after n failures. This discussion is summarized with a proposition and illustrated by Figure 4.

Proposition 3. *Under (4) and (35), the promotion rungs and the demotion rungs of any job $k = 1, \dots, K - 1$ form two partitions of the interval $[\rho_k, \rho_{k+1})$ as in (36) and (37). A worker in promotion rung $[\bar{q}_{kn}, \bar{q}_{kn-1})$ is promoted to job $k + 1$ after n successes. A worker in demotion rung $[\underline{q}_{kn}, \underline{q}_{kn+1})$ is demoted to job $k - 1$ after n failures.*

3.2.3 Promotions and Demotions by Multiple Levels

So far I have focused on the number of successes it takes to be promoted by exactly one level and on the number of failures it takes to be demoted by exactly one level, under the assumption that all promotions and demotions are at most by one level. More generally, it is possible that in certain regions of the interval $[\rho_k, \rho_{k+1})$, a single success could lead to a promotion by one level, a promotion by two levels, and so on. Here I focus on partitioning this interval $[\rho_k, \rho_{k+1})$ into subintervals by the level of job assigned after one success. The case for demotions by one or more levels is analogous.

To help develop this general characterization, it is useful to define the inverse functions $q_H(p)$ of (6) and $q_L(p)$ of (7) by

$$q_H(p) = P_H^{-1}(p) = \frac{\beta p}{\alpha - (\alpha - \beta)p}, \quad (38)$$

$$q_L(p) = P_L^{-1}(p) = \frac{(1 - \beta)p}{1 - \alpha + (\alpha - \beta)p}. \quad (39)$$

The interpretation of $q_H(p)$ is that a worker who starts with prior $q_H(p)$ and succeeds in job k ends with prior p , whereas the interpretation of $q_L(p)$ is that a worker who starts with prior $q_L(p)$ and fails in job k ends with prior p .

These inverse functions are used to represent the rule that specifies the job assigned after a success or a failure in a given job k as a partition of the interval $[\rho_k, \rho_{k+1})$. First consider what happens after a success at the lower edge ρ_k and at the upper edge ρ_{k+1} . At prior ρ_k , a success leads to job $k_H(\rho_k) \geq k$, whereas at prior ρ_{k+1} a success leads to job $k_H(\rho_{k+1}) \geq k + 1$.

Now, as the prior increases from ρ_k , at first the worker will continue to be assigned to the same job $k_H(\rho_k)$ until (by continuity of P_H) a cutoff prior is reached, at which one success leads the prior to increase exactly to $\rho_{k_H(\rho_k)+1}$, the lowest prior at which job $k_H(\rho_k) + 1$ is optimal. This cutoff prior is the inverse image of $\rho_{k_H(\rho_k)+1}$ under P_H , which is denoted by $q_H(\rho_{k_H(\rho_k)+1})$. This part of the construction is illustrated by Figure 5.

Continuing in this fashion, a partition of $[\rho_k, \rho_{k+1})$ can be constructed. That is, after succeeding in job k , a worker with a prior in $[\rho_k, q_H(\rho_{k_H(\rho_k)+1}))$ is assigned to job $k_H(\rho_k)$, a worker with a prior in $[q_H(\rho_{k_H(\rho_k)+1}), q_H(\rho_{k_H(\rho_k)+2}))$ is assigned to job $k_H(\rho_k) + 1$, and so on, until the uppermost part of $[\rho_k, \rho_{k+1})$ is reached, namely, $[q_H(\rho_{k_H(\rho_{k+1})}), \rho_{k+1})$, where a success leads to the assignment of $k_H(\rho_{k+1})$. The partition of $[\rho_k, \rho_{k+1})$ after a failure in job k is analogous.

Notice, of course, that if job k is not very informative, there may not be much variation in posterior beliefs and, therefore, no significant change in the job assigned from period t to period $t + 1$. For example, if job k is completely uninformative, both of these subdivisions collapse in that $k_L(p) = k_H(p) = k$ at all $p \in [\rho_k, \rho_{k+1})$. Regardless of success or failure in job k , job k is still assigned.

3.2.4 Symmetric versus Asymmetric Signals

Note that, as mentioned, the informational structure here allows for asymmetric signals. The case of symmetric signals is simpler analytically, but I argue that it is less relevant empirically. With symmetric signals at any job k , that is, with $\beta = 1 - \alpha$, it follows $P_H(P_L(p)) = P_L(P_H(p)) = p$. Thus, in a precise sense, one failure exactly offsets one success. This property implies that $\bar{n}_k = \underline{n}_k$ and both equal, say, n_k . That is, if it takes n_k successes to be promoted to job $k + 1$ starting from ρ_k , then it takes exactly the same number of failures to be demoted to job $k - 1$ starting from (just under) ρ_{k+1} . Moreover, if a worker is promoted from job k to job $k + 1$ after a success, then if that newly promoted worker fails at job $k + 1$, that worker is necessarily demoted back to job k .

Asymmetric signals, that is, cases in which $\beta \neq 1 - \alpha$, allow for much more flexible patterns of promotions and demotions. For example, consider an extreme case in which $\alpha = 1$ so that high-ability workers never fail and β is close to one so that low-ability workers rarely fail. In this case, a success is not very informative. Therefore, it takes a large number of successes for the prior to increase past any given cutoff for promotion. Also, demotions are rare. This combination of patterns is commonly observed in the data. Notice that this case is very different from the symmetric case, in which if high-ability workers always succeed ($\alpha = 1$), then low-ability workers always fail ($\beta = 1 - \alpha = 0$). In the data, the general pattern of gradual promotions and rare demotions is typical. In the BGH data, in Pastorino (2013) I find support for the asymmetric signal case in which β is close to α .

3.3 Turnover

This characterization of job assignment and wages applies to any order $l = l(f, k)$ of the \bar{K} jobs of the economy. Since firms are collections of such jobs, the model's implications for the mobility of workers across firms depends only on how these jobs are distributed across firms.

To illustrate how the interval characterization applies to a given distribution of jobs across firms, let there be three firms A , B , and C . Let firm A have jobs $1, \dots, K_A$, firm B have jobs $1, \dots, K_B$, and firm C have jobs $1, \dots, K_C$. Suppose they satisfy (21) with the order $l = l(f, k)$. Suppose the jobs of firm A happen to be the lowest-ranked ones and the jobs of firm C the highest-ranked ones. Then, the ordering of jobs and firms is

$$1, \dots, K_A; 1, \dots, K_B; 1, \dots, K_C,$$

where $l = 1$ corresponds to the first job of A and $l = K_A + K_B + K_C$ corresponds to the last job of K_C . The interval characterization of job assignment then implies a hierarchy of firms as well as a hierarchy of jobs within each firm. To see this, suppose that a worker starts at job k of firm B . Continued success leads the worker to being first promoted to higher-ranked jobs at firm B , then to turn over to the higher-ranked firm C and, next, to work up the hierarchy of jobs at firm C . Likewise, continued failure by this worker leads to demotion first to jobs in the hierarchy of firm B , eventually to turnover to the lower-ranked firm A , and then to demotion to the lower-ranked jobs of firm A 's hierarchy.

Another situation is one in which each job at level k of a firm corresponds to a different type of job. For instance, job 1 of each of the three firms may be an administrative professional job. Suppose that jobs are ranked as follows: job 1 of firm A , then job 1 of firm B , then job 1 of firm C . Then, a worker who succeeds will be promoted up the ranks of the administrative professional jobs by moving to the administrative professional jobs of firm A , then B , then C .

Note that as long as the ordering (4) applies to the jobs of each firm and (21) applies to all jobs, success may lead a worker to crisscross between firms on the way up the $l(f, k)$ ladder.

4 Mixed Ability

Here, I relax the assumption that ability is general across all the firms in the economy. I am motivated to do so by two sets of findings. Starting from Jovanovic (1979), the argument has long been made that wages increase with tenure in a firm as workers accumulate firm-specific capital in the form of information about the match of their skills to a firm's jobs. For example, Flinn (1986) and Topel (1991) document that firm-specific capital can explain why the hazard of separation from a firm eventually decreases with tenure in the firm as well as the dynamics

of wages with tenure in a firm.

A more recent strand of the literature has emphasized the importance of occupation-specific capital in order to explain job turnover and wage growth. Miller (1984) is among the first papers to provide evidence that job turnover patterns can be interpreted as due to the accumulation of information about the match of a worker’s skill to jobs, when workers have different priors about their suitability to jobs in different occupations. Moscarini and Vella (2008) document that occupational mobility decreases with experience, and, more recently, Kambourov and Manovskii (2009a,b) estimate that returns to occupational tenure are substantial.⁷

In terms of the first literature, my model of general ability does not naturally lead to hazards of separation from a firm that decrease with tenure, nor does it explain why workers would consistently suffer wage losses upon separation. In terms of the second literature, even if an occupation is interpreted as a collection of jobs in the general setup, the model is silent on the patterns of the hazard of job separation and wages for occupations.

I consider two extensions of the model that give rise to either firm-specific or occupation-specific capital, produce patterns consistent with the evidence just discussed, and still account for a hierarchy of jobs in each firm.

4.1 Firm-Specific Ability

Jovanovic’s (1979) work on turnover assumes that each firm has a single job and that a worker’s ability is firm specific in that each worker independently draws the quality of the match with each firm. My model of firm-specific ability generalizes Jovanovic’s model to an environment of multi-job firms by letting abilities be firm specific but general across a firm’s jobs.

Specifically, let the worker independently draw a skill type $s_f \in \{a, b\}$ for each firm $f = 1, \dots, N$. Then a worker’s ability is a vector $s = (s_1, \dots, s_N)$ and a worker’s prior is a vector $p = (p_1, \dots, p_N)$. Assume that output at the jobs of firm f depends only on the component s_f of a worker’s skill and that these jobs are ordered by (4). Notice here that a worker’s ability at a given firm does not provide information about the ability of the worker at any other firm.

To set up the optimal job assignment problem, let $y_f(p_f) = \max_{k \in K_f} y_f(p_f, k)$. Treating $y_f(p_f)$ as the expected output of firm f , the job assignment problem consists of two parts: to assign a worker to a firm and then to assign the worker to the job with the highest expected output in that firm. In this setting, the assignment of workers to firms is a standard *independent* multi-armed bandit problem for which the solution takes a simple form: assign each firm an index, $M_f(p_f)$, called the Gittins index, which is based on a particular (artificial) programming problem. Then, in any period, given the vector of beliefs $p = (p_1, \dots, p_N)$, choose the firm

⁷See Neal (1995) and Parent (2000) for the related notion of industry-specific capital.

with the highest index. The programming problem that defines the index for firm f is the following. At each point in time, workers at firm f and in the current period receive the period payoff $y_f(p)$ or stop permanently and receive a terminal reward of m . The Gittins index is the minimum amount of terminal reward m that leads to indifference between stopping today and receiving m or continuing employment and facing the same decision in the next period. Hence,

$$M_f(p_f) = \min \{m | J_f(p_f, m) = m\}, \quad (40)$$

where

$$J_f(p_f, m) = \max\{m, (1 - \delta)y_f(p_f) + \delta E J_f(p'_f, m | p_f)\} \quad (41)$$

and the expectation in (41) is defined based on (6) and (7). The N problems of the form (41), one for each firm, should be thought of as convenient devices used in the definition of the optimal policy and have no other interpretation. (See, for example, Banks and Sundaram (1992) for details.)

Clearly, in this case, if firm f is chosen at t and success occurs, then firm f will again be chosen at $t + 1$. To see why, note that if firm f is chosen at t given the vector of beliefs p , by definition $M_f(p_f) \geq M_g(p_g)$ for all $g \neq f$. Now, a success at firm f leads to a new vector of beliefs p' , where $p'_f = P_H(p_f)$ and the other elements are unchanged. Since J_f , and hence M_f , are increasing in p_f , if employment at firm f is optimal at p , employment is also optimal at p' .

Hence, the only way a worker can turn over to a new firm is for a worker to experience sufficiently many failures so that the modified Gittins index of firm f falls below the Gittins index of another firm (evaluated at the prior at entry into firm f). A successful worker at a given firm f , instead, will advance to higher-level jobs within that firm. This discussion is summarized in the following proposition.

Proposition 4. *Under (4) with firm-specific ability that is general across a firm's jobs, the optimal assignment rule follows a modified Gittins index policy of the form (40) and (41). Under this policy, success in firm f leads to promotion in that firm according to the hierarchy of jobs in that firm. Failure leads to demotion within that firm according to that hierarchy, then eventually to switching to another firm.*

This model will clearly imply the two observations I discussed to motivate it. The hazard rate of separation will eventually decrease with tenure because of selection: over time, the workers who stay with a given firm are more and more likely to be of high ability (for that firm), since the workers with low ability gradually separate. Indeed, *only* such workers separate. Thus, the workers who stay with a firm will, on average, have a high probability of success and experience wage increases and promotions.

4.2 Occupation-Specific Ability

Here, extending Miller (1984), I allow ability to be occupation specific but general across firms. To make ideas more concrete, suppose that there are three firms, A , B , and C , and two occupations, $O \in \{I, R\}$, say, information technology and human resources, respectively. An occupation is interpreted as a collection of jobs for which a given aspect of a worker's skills are most relevant. I formalize this idea by letting a worker be endowed with two sets of skills, each of which is specific to a certain group of jobs termed an *occupation*. The similarity of the skill requirements of the jobs of an occupation can be thought of as capturing a measure along which jobs are close in a technological sense.

Let the occupation skill vector of each worker be $s = (s_I, s_R)$ with $s_I, s_R \in \{a, b\}$. Let $p = (p_I, p_R)$ denote the priors that the worker is high skilled in occupations I and R . Let each firm f have K_I jobs that are complementary to skill I with expected output $y_f(p_I, k)$ for $k = 1, \dots, K_I$ and K_R jobs that are complementary to skill R with expected output $y_f(p_R, k)$ for $k = K_{I+1}, \dots, K_{I+R}$.

Now consider two sets of jobs, the $3K_I$ jobs of occupation I and the $3K_R$ jobs of occupation R . The results immediately extend to this setup as long as these two sets of jobs can be separately ordered. Suppose then that there exist two orderings: an ordering $l = l_I(f, k)$ that orders the $3K_I$ jobs of occupation I according to (21) and an ordering $l = l_R(f, k)$ that orders the $3K_R$ jobs of occupation R according to (21).

In this case, the optimal assignment of jobs within each occupation has an interval characterization under the order $l_O(f, k)$ for $O \in \{I, R\}$, and the optimal policy across occupations has a modified Gittins index form. Specifically, the optimal policy has a Gittins index form similar to (40),

$$M_O(p_O) = \min \{m \mid J_O(p_O, m) = m\}, \quad (42)$$

and for occupation O , (41) is now given by

$$J_O(p_O, m) = \max \{m, (1 - \delta) \max_{k \in K(O)} y_f(p_O, k) + \delta E J_O(p'_O, m \mid p_O)\}, \quad (43)$$

where $K(I)$ is the set of jobs 1 to K_I in each of the three firms and $K(R)$ is the set of jobs K_{I+1} to K_{I+R} in each of the three firms. The optimal policy is to choose the occupation O using the Gittins index and then choose the job within that occupation using the interval characterization of the jobs $K(O)$.

I have used a simple example with two occupations and three firms to make the ideas concrete, but the following proposition clearly applies to an arbitrary number of firms and occupations.

Proposition 5. *Under (4) applying separately to each occupation O and occupation-specific ability that is general across firms, the optimal assignment rule follows a modified Gittins index policy of the form (42) and (43). Under this policy, success in occupation O leads to promotion in that occupation according to the hierarchy $l_O(f, k)$ of jobs in that occupation. Failure leads to demotion within that occupation according to the hierarchy $l_O(f, k)$, then eventually to switching to another occupation.*

To interpret this proposition, note that the hierarchy of jobs in an occupation is given by the ordering of $l_O(f, k)$. Depending on how the jobs in this hierarchy are distributed across firms, moving up the hierarchy may involve mostly moving up within firms, then crossing over to higher-ranked firms within an occupation (in the sense that some firms have a low ranking of their jobs by $l_O(f, k)$, whereas others may have a high ranking), or it may involve moving up by crisscrossing back and forth between firms (if firms contain a mix of low-, medium-, and high-ranked jobs according to this order). Clearly, this model produces the patterns described above: a hazard of switching occupations that is (eventually) decreasing with job tenure or experience, and wage rates that tend to rise with tenure in an occupation.

Some recent work has argued that these patterns are not universal. In particular, using Danish data, Groes, Kircher, and Manovskii (2010) document that occupational mobility is U-shaped in that both the relatively low and the relatively high wage earners within an occupation have a particularly high probability of leaving the occupation. The low wage earners tend to switch to new occupations with lower average wages, and the high wage earners tend to switch to new occupations with higher average wages.

A version of the model can easily generate these facts. To see how, consider the following setup. As above, let a worker have a set of skills $s = (s_I, s_R)$. Let there be two sets of occupations $\{I_1, I_2, \dots, I_{N_I}\}$ and $\{R_1, R_2, \dots, R_{N_R}\}$, and suppose that ability is independent across the jobs associated with the two sets but general across the jobs within each set. Specifically, suppose that the productivity of jobs in occupations I_1, I_2, \dots, I_{N_I} depends only on s_I , whereas the productivity of jobs in occupations R_1, R_2, \dots, R_{N_R} depends only on s_R . Let the jobs in each set of occupations be ranked by (21). Let this ordering also be such that occupation I_1 has jobs ranked lower than those of I_2 , and so on. Specifically, let the order $l_I(f, k)$ rank the jobs of occupations I_1, I_2, \dots, I_{N_I} as follows: the jobs of I_1 run from $l = 1$ to $l = K_{I_1}$, the jobs of I_2 run from $l = K_{I_1+1}$ to $l = K_{I_1+I_2}$, and so on. Similarly, let the order $l_R(f, k)$ rank the jobs in occupations R_1, R_2, \dots, R_{N_R} . Then, Proposition 5 applies with the occupations $\{I_1, I_2, \dots, I_{N_I}\}$ treated as one grand (artificial) occupation with $K_I = \sum_{i=1}^{N_I} K_{I_i}$ jobs, and the occupations $\{R_1, R_2, \dots, R_{N_R}\}$ as treated as a second grand (artificial) occupation with $K_R = \sum_{i=1}^{N_R} K_{R_i}$ jobs.

Given this ordering, consider a worker at the lowest-ranked job in an occupation, say, job

$K_{I_1} + 1$ of occupation I_2 . That worker will have the lowest wage in that occupation, and a single failure will lead to a “demotion” to the next lower occupation I_1 with lower wages. Next consider a worker at the highest-ranked job in an occupation, say, job $K_{I_1} + K_{I_2}$ of occupation I_2 . That worker will have the highest wage in occupation I_2 , and a single success will lead to a “promotion” to the next higher occupation I_3 with higher wages. Workers in medium-ranked jobs in an occupation will be less likely to switch occupations. Thus, this setup can easily explain how the U-shaped pattern of switching between occupations may arise. Finally, workers will switch between the two sets of occupations only when the prior about the skill relevant to that set of occupations drops so low that the associated modified Gittins index for that set is lower than the modified Gittins index for the other set.

5 Experimentation

Here, experimentation is allowed for by having jobs differ in their informativeness. I start by having all firms identical but some jobs possibly more informative than others, then discuss how the results apply to the case of heterogeneous firms.

5.1 The Multi-Armed Bandit Problem

Since all firms are identical, I drop the f subscript in (20) and write the planning problem as

$$W(p_t) = \max_{k \in K} [(1 - \delta)y(p_t, k) + \delta EW(p_{t+1}|p_t, \theta_k)]. \quad (44)$$

This problem is an instance of a *multi-armed bandit* with dependent arms. In this problem, the planner values jobs according to both their static expected output and their informativeness. Specifically, consider the relative benefits of any two jobs k and k' . This benefit can be decomposed into two terms: the *static gain* in expected output from job k relative to job k' , namely,

$$y(p_t, k) - y(p_t, k') \quad (45)$$

and the *information gain* of job k relative to job k' , namely,

$$EW(p_{t+1}|p_t, \theta_k) - EW(p_{t+1}|p_t, \theta_{k'}). \quad (46)$$

Now, job k is preferred to job k' if, after applying discounting, the static gain of job k relative to k' is larger than the information gain of job k' relative to job k . Of course, if job k has both a static gain and an information gain relative to job k' , then it is obviously preferred to job k' .

The information benefit of one job over another arises from the convexity of the value

function $W(p)$. It is straightforward to prove that the value function $W(\cdot)$ is convex in p . (See, for instance, Banks and Sundaram (1992) for the independent case. Adapting their proof to this case is easy.) This convexity implies that mean-preserving spreads of posterior beliefs are beneficial. More precisely, when job k is more informative than job k' , defined as in (8), and the value function is convex, then using (6) and (7) we have that (46) is positive.

The intuitive idea is that the more spread out the posterior, the greater the information acquired, and thus, the more precisely the newly assigned job can be tailored to the state. That the greater informativeness of a job leads to that job having greater information benefits follows from the convexity of the value function, as illustrated by Figure 6.

It is well known that, in contrast to the independent arm case, no general characterization result of the optimal policy for (44) exists. Here, I will characterize the solution to (44) for some special cases that I argue are empirically relevant.

5.2 Allocations in an Entry Job Economy

A standard intuition in the literature on careers in firms, dating back to Holmström and Tirole (1989), is that certain jobs in firms, especially entry-level ones, may have greater information content than others, so they may be used by a firm as a screening device to acquire information about the ability of workers. (See also Felli and Harris (2004) for a similar interpretation.) As mentioned, in Pastorino (2013) I document that the lowest-level job is the most informative and the rest of the jobs are essentially equally informative.

Motivated by this intuition and these empirical findings, I suppose that job 1, called the *entry job*, is more informative than jobs $2, \dots, K$, each of which is equally informative. Specifically, I assume that the entry job has informativeness $\theta_1 = (\alpha_1, \beta_1)$, whereas jobs $2, \dots, K$ are less informative in that their informativeness is $\theta = (\alpha, \beta)$, with

$$\frac{\alpha_1}{\beta_1} > \frac{\alpha}{\beta} \text{ and } \frac{1 - \alpha_1}{1 - \beta_1} < \frac{1 - \alpha}{1 - \beta}. \quad (47)$$

Assume also that the K jobs satisfy (4) and (27). Let $\{\rho_k\}_{k=2}^{K-1}$ denote the corresponding static cutoffs. Under these conditions, the solution to (44) has the following form: expand the static interval $[0, \rho_2)$ of priors at which job 1 is assigned to $[0, \phi_{k^*})$ with $\phi_{k^*} > \rho_2$; at all priors p larger than ϕ_{k^*} , assign the job that would have been assigned under the static assignment rule. Note that the interval of priors at which job 1 is assigned can expand to the point where job 1 “crowds out” some jobs starting from job 2 on. Let k^* denote the lowest-level job that is not crowded out. Then, the static assignment rule applies starting with job k^* at priors in the interval $[\phi_{k^*}, \rho_{k^*+1})$. Note that job 1 cannot completely crowd out the highest job K : as the prior p approaches 1, the information gain of job 1 goes to zero. Hence, $\phi_{k^*} < 1$. If job 1

partially crowds out job K , then job K is used only at priors in $[\phi_{k^*}, 1]$, a strict subinterval of $[\rho_K, 1]$; otherwise, job K is assigned at priors in the entire interval $[\rho_K, 1]$.

More formally, I claim that the solution to the entry job problem has an interval characterization of the form

$$[0, \phi_{k^*}), [\phi_{k^*}, \rho_{k^*+1}), [\rho_{k^*+1}, \rho_{k^*+2}), \dots, [\phi_K, 1] \quad (48)$$

in that there exist dynamic cutoffs ϕ_{k^*} and ϕ_K such that for priors in $[0, \phi_{k^*})$ the entry job is assigned. Let k^* be defined as the smallest job k such that the upper edge of the interval $[\rho_k, \rho_{k+1})$ for which job k would be statically assigned, namely, ρ_{k+1} , is strictly greater than the dynamic cutoff ϕ_{k^*} . Define

$$\phi_K = \left\{ \begin{array}{l} \rho_K \text{ if } \phi_{k^*} < \rho_K \\ \phi_{k^*} \text{ otherwise} \end{array} \right\}.$$

Note that when $\phi_K = \rho_K$, job K is assigned over the static interval $[\rho_K, 1]$. When $\phi_{k^*} \geq \rho_K$, job K is assigned only at priors in $[\phi_{k^*}, 1]$. See Figure 7.

In sum, the job assignment rule associated with (48) is as follows: at $p \in [0, \phi_{k^*})$ assign the entry job, and at $p \in [\phi_{k^*}, 1]$ assign the job that would have been assigned using the static assignment rule. I also claim that ϕ_{k^*} satisfies

$$\rho_2 < \phi_{k^*} < 1. \quad (49)$$

Proposition 6. *In the entry job economy, if job 1 is sufficiently more informative than the other jobs, and (4) and (27) hold, then there exists ϕ_{k^*} satisfying (49) such that the optimal job assignment rule is dynamic and is given by the interval characterization (48).*

Proof: Since jobs $k = 2, \dots, K$ have the same informativeness, we can write the value $W(p_t)$ in (44) as the choice between using job 1 today or using the statically best job among jobs $2, \dots, K$ so that $W(p_t) = \max \{W_1(p_t), \bar{W}(p_t)\}$, where

$$W_1(p_t) = (1 - \delta)y(p_t, 1) + \delta EW(p_{t+1}|p_t, \theta_1) \quad (50)$$

$$\bar{W}(p_t) = (1 - \delta) \max_{k=2, \dots, K} y(p, k) + \delta EW(p_{t+1}|p_t, \theta). \quad (51)$$

Suppose first that job 1 is perfectly informative so that $\alpha_1 = 1$ and $\beta_1 = 0$. Then,

$$EW(p_{t+1}|p_t, \theta_1) = p_t W(1) + (1 - p_t)W(0) = p_t y(1, K) + (1 - p_t)y(0, 1),$$

where I have used that $r(p_t, \theta_1) = p_t$, that $P_{H1}(p_t) = 1$ and $P_{L1}(p_t) = 0$ at $p_t \in (0, 1)$, that priors of 0 and 1 are absorbing states, and, finally, that under (21), job 1 is the best assignment

at $p_t = 0$ and job K is the best assignment at $p_t = 1$. In particular,

$$W_1(p_t) = (1 - \delta)y(p_t, 1) + \delta [p_t y(1, K) + (1 - p_t)y(0, 1)]. \quad (52)$$

Note that by (21) $W_1(0) = y(0, 1) > \bar{W}(0) = y(0, 2)$ and $W_1(1) = (1 - \delta)y(1, 1) + \delta y(1, K) < \bar{W}(1) = y(1, K)$. Moreover, as is apparent from (52), $W_1(p_t)$ is a straight line, whereas by (21) and the convexity of $W(p_t)$, $\bar{W}(p_t)$ is an increasing convex function of p_t . Therefore, $W_1(p_t)$ crosses $\bar{W}(p_t)$ once from below to above at ϕ_{k^*} as depicted by Figure 8.

Next, I claim that $\phi_{k^*} > \rho_2$. First I show that $\phi_{k^*} \geq \rho_2$. To see why, suppose by way of contradiction that $\phi_{k^*} < \rho_2$, that is, $W_1(p) = \bar{W}(p)$ at $p = \phi_{k^*} < \rho_2$. By single-crossing of $W_1(p_t)$ and $\bar{W}(p_t)$, this implies that there exists a $p \in (\phi_{k^*}, \rho_2)$ such that $W_1(p) < \bar{W}(p)$. This is impossible, since by (21) and the definition of the static cutoff ρ_2 , $y(p, 1) > y(p, 2)$ at any $p < \rho_2$. Moreover, at all priors p_t , by convexity $EW(p_{t+1}|p_t, \theta_1) \geq EW(p_{t+1}|p_t, \theta)$, since job 1 is more informative than the other jobs. Hence, for $p \in (\phi_{k^*}, \rho_2)$, $W_1(p) > \bar{W}(p)$, which is a contradiction. Second, ϕ_{k^*} cannot equal ρ_2 . The reason is that since ρ_2 is interior and job 1 is strictly more informative than the other jobs, job 1 has a strict information gain over other jobs in that $EW(p_{t+1}|\rho_2, \theta_1) - EW(p_{t+1}|\rho_2, \theta) > 0$. In addition, job 1 is tied statically with job 2. Hence, $W_1(\rho_2) > \bar{W}(\rho_2)$.

Next, note that by construction of \bar{W} , at any p such that $\bar{W}(p) > W_1(p)$, the job assigned is given by $k^*(p) = \arg \max_{k=2, \dots, K} y(p, k)$, which is the statically optimal job. Finally, that the same result holds when the entry-level job is sufficiently informative follows by continuity, for α_1 sufficiently close to 1 and β_1 sufficiently close to zero. \square

I think of this result as formalizing the intuition of Holmström and Tirole (1989). Note that the dynamic job assignment rule (48) in the entry job economy also has implications for wages. To do so, consider decentralizing the economy using the wage function (15) discussed earlier. In particular, let $w(p, \theta_1) = y(p, 1)$ and $w(p, \theta) = \max_{k=2, \dots, K} y(p, k)$.

Corollary 4. *In an entry job economy at priors in (ρ_2, ϕ_{k^*}) , workers accept a negative compensating differential in the entry job in that when assigned to that job, they receive the wage $w(p, \theta_1) < w(p, \theta)$. As p crosses the cutoff ϕ_{k^*} , wages increase discretely by the amount*

$$w(\phi_{k^*}, \theta) - w(\phi_{k^*}, \theta_1) = \frac{\delta}{1 - \delta} [EW(p_{t+1}|\phi_{k^*}, \theta_1) - EW(p_{t+1}|\phi_{k^*}, \theta)]. \quad (53)$$

Proof: I first show that wages discretely increase at the cutoff. To do so, note that by the definition of the cutoff ϕ_{k^*} , $W_1(\phi_{k^*}) = \bar{W}(\phi_{k^*})$ so that using the definitions of $w(p, \theta_1)$ and $w(p, \theta)$ in (50) and (51) yields (53). Next, since job 1 is more informative than the other jobs, given any (interior) prior, the posterior under job 1 is a mean-preserving spread of that under jobs $2, \dots, K$. Since the value function $W(p)$ is convex in p , it follows that the right side of

(53) is positive. Next, to establish the first part of the corollary, note that since $\phi_{k^*} > \rho_2$, (21) implies that for any $\rho_2 < p \leq \phi_{k^*}$, $y(p, 1)$ is strictly smaller than $y(p, 2)$. At priors in the interval $(\rho_2, \phi_{k^*}]$, the difference $y(p, 2) - y(p, 1)$ is largest at ϕ_{k^*} . \square

Notice that the corollary offers a novel interpretation of the wage premium that workers typically experience at promotion and, therefore, of the forces that determine wage growth with tenure in a firm. Here, a worker with a prior slightly smaller than ϕ_{k^*} is assigned to a job at which the worker has a static comparative disadvantage at that prior by (21). Such an assignment is nonetheless efficient because the information benefit to the worker from working at job 1, in the form of higher expected future wages, more than offset the current loss in wages. As a result, the lower current wage in job 1 contains a negative *compensating differential* for the greater information value of job 1 relative to job k^* . Indeed, as (53) makes clear, the jump in the wages at ϕ_{k^*} equals the information gain of job 1 over the other jobs.

So far I have presumed that the lowest-level job, job 1, is more informative than the remaining jobs. For completeness, consider in contrast a *top job economy* in which the highest-level job, job K , is more informative than all the other jobs. Clearly, a result analogous to Proposition 6 holds. Specifically, relative to the equally informative case, the interval at which job K is assigned increases to the left, implying that job K potentially crowds out jobs $K - 1$, $K - 2$, and so on. That is, workers are assigned to job K at priors in the interval $[\phi_K, 1]$ with $\phi_K < \rho_K$, whereas to the left of ϕ_K the static assignment rule applies. Let h^* denote the job that is assigned immediately to the left of ϕ_K . See Figure 9.

Observe also that in a top job economy, logic analogous to that used in Corollary 4 implies that wages at promotion decrease by the discrete amount $w(\phi_K, \theta_{h^*}) - w(\phi_K, \theta_K)$. The reason is that now as the prior crosses the cutoff ϕ_K from below, workers switch from being assigned to job h^* to job K . In the interval $[\rho_{h^*}, \phi_K)$, by definition $y(p, h^*) > y(p, K)$. But job K provides an information benefit relative to job h^* . Hence, job h^* contains a positive compensating differential relative to job K in the interval $[\rho_{h^*}, \phi_K)$ in that

$$w(\phi_K, \theta_{h^*}) - w(\phi_K, \theta_K) = \frac{\delta}{1 - \delta} [EW(p_{t+1} | \phi_K, \theta_K) - EW(p_{t+1} | \phi_K, \theta_{h^*})] < 0.$$

Here, a worker's wage will drop discretely as the prior crosses the threshold ϕ_K .

In light of these features, I think that the implications of the entry job economy are much more in line with the data than those of the top job economy. Using the lens of this model, the observation that wages jump upon promotion is supportive of the idea that lower-level jobs are more informative than higher-level jobs. Hence, the greater informativeness of lower-level jobs could be an important factor in explaining the wage increases typically observed at promotions. (In Pastorino (2013) I provide evidence for these patterns by estimating a general version of this model using the BGH data.) This result could also explain why, at least in the BGH data, even

for workers with high tenures and experience who are likely to have extreme priors, promotions lead to large wage increases. See the discussion in Gibbons and Waldman (2006) about the inability of their model to generate this pattern.

Finally, Proposition 6 can be used to illustrate how the relative informativeness of the entry-level job relative to the non-entry jobs affects the speed of promotion. To this end, let economy 1 have an entry-level job of a given informativeness θ_1 and the non-entry-level jobs have informativeness θ . Let economy 2 have the same entry-level job, but suppose that the non-entry-level jobs have lower informativeness θ' . Then the dynamic cutoff ϕ_{k^*} for economy 1 is smaller than the corresponding cutoff for economy 2. But the updating rules $P_H(p, \theta_1)$ and $P_L(p, \theta_1)$ are the same in the two economies. Consider now a worker with a prior p lower than ϕ_{k^*} . It immediately follows that the average time until the first promotion for the worker is shorter in economy 1 than in economy 2.

Corollary 5. *In an entry job economy, as the level of informativeness of the non-entry-level jobs decreases, the average time until first promotion out of the entry-level job increases.*

5.3 Decentralization

So far I have considered only the planning problem and have been silent on how these jobs are distributed across firms. It should be clear and intuitive that given any collection of \bar{K} of the jobs in the economy that satisfy (21), the implications for careers within firms and turnover between firms differ greatly depending on how these jobs are distributed.

To see this, suppose that in the economy there are entry-level jobs that are very informative and non-entry-level jobs that are equally informative. Consider a distribution such that technology K_A consists of only entry-level jobs and the other technologies, say, technologies K_B and K_C , have only non-entry-level jobs. Then, the problem of each firm is static and all experimentation takes place for workers turning over across firms. Alternatively, suppose that the technologies of all firms are similar in that each firm has a mixture of entry-level jobs and non-entry-level jobs. Then, each firm faces a dynamic allocation problem, so experimentation also occurs within firms. In the first case, a firm's allocation policy is static, but a worker's choice of firm dynamic. In the second case, both a firm's allocation policy and a worker's choice of firm are dynamic.

5.4 An Example

Here, motivated by my findings in Pastorino (2013), I present a simple two-period example to further illustrate the dynamics of jobs and wages in equilibrium for an entry job economy. In my empirical work based on the BGH data, I find that the lowest of the three main levels of

the firm's hierarchy is the most informative. In my sample, each promotion is by one step and no demotion is observed. Here, I construct an example that displays these features.

Let job 1 be the most informative, job 2 the next most, and job 3 the least informative. Assume that these jobs satisfy (21). I proceed by constructing equilibrium and then deriving sufficient conditions for this equilibrium to be unique. Figure 10 illustrates the basic idea. In the second period, since there is no benefit to more information, the allocations are given by the static cutoffs: work in jobs 1, 2, and 3 as the prior ranges from $[0, \rho_2)$ to $[\rho_2, \rho_3)$ to $[\rho_3, 1]$. In the first period, the allocations are given by the dynamic cutoffs: work in jobs 1, 2, and 3 as the prior ranges from $[0, \phi_2)$ to $[\phi_2, \phi_3)$ to $[\phi_3, 1]$. Note that since job 1 is more informative than job 2, the dynamic cutoff ϕ_2 is to the right of the static cutoff ρ_2 , so $\phi_2 > \rho_2$. Similarly, since job 2 is more informative than job 3, $\phi_3 > \rho_3$.

I claim that under three conditions, this dynamic path of jobs assignments is optimal and entails promotions by one step but no demotions. The conditions amount to parameter restrictions. The first two conditions are

$$P_{L2}(\phi_2) \geq \rho_2 \text{ and } P_{L3}(\phi_3) \geq \rho_3 \quad (54)$$

and rule out demotions. The first inequality implies that even if a worker fails at the lowest prior at which job 2 is assigned, namely ϕ_2 , then the resulting prior $P_{L2}(\phi_2)$ is still above the static cutoff for job 2, ρ_2 . Hence, all workers assigned to job 2 in the first period that fail will still be assigned to job 2 in the second period. The second inequality analogously rules out demotions from job 3. The third condition,

$$P_{H1}(\phi_2) < \rho_3, \quad (55)$$

rules out promotions by more than one level. This inequality requires that even if a worker succeeds at the highest prior at which job 1 is assigned in the first period, then the resulting prior $P_{H1}(\phi_2)$ is still below the static cutoff for job 3, ρ_3 .

Under these restrictions, the conditions defining the dynamic cutoff ϕ_2 are $W_1(\phi_2) = W_2(\phi_2)$, where $W_1(\phi_2)$ equals

$$(1 - \delta)y(\phi_2, 1) + \delta [r(\phi_2, \theta_1)y(P_{H1}(\phi_2), 2) + (1 - r(\phi_2, \theta_1))y(P_{L1}(\phi_2), 1)] \quad (56)$$

and $W_2(\phi_2) = y(\phi_2, 2)$. To understand $W_1(\phi_2)$, note that under the plan associated with (56), in period 1 the worker works at job 1 at prior ϕ_2 . After a success, the prior is updated to $P_{H1}(\phi_2)$ and (55) implies that job 2 is optimal at that prior.

The only subtlety is the argument for the job assignment after a failure in job 1 at ϕ_2 . After

a failure, the prior is updated to $P_{L1}(\phi_2)$, and I argue that the first no demotion condition in (54) implies that the worker is then assigned to job 1. By way of contradiction, suppose instead that the worker is assigned to job 2. But this would imply that a worker who starts in job 1 at ϕ_2 works in job 2 in the second period regardless of success or failure. In turn, this would imply that there is no value to more information at ϕ_2 , so $\phi_2 = \rho_2$, which is inconsistent with the first no demotion condition. To understand $W_2(\phi_2)$, note first that since job 1 is more informative than job 2, $P_{H2}(\phi_2) \leq P_{H1}(\phi_2)$, which by (55) implies that after a success, job 2 is optimal. Next, note that after a failure in job 2, the prior is updated to $P_{L2}(\phi_2)$. By the first inequality in (54), job 2 is optimal.

The conditions determining the dynamic cutoff ϕ_3 are $W_2(\phi_3) = W_3(\phi_3)$, where $W_2(\phi_3)$ equals

$$(1 - \delta)y(\phi_3, 2) + \delta [r(\phi_3, \theta_2)y(P_{H2}(\phi_3), 3) + (1 - r(\phi_3, \theta_2))y(P_{L2}(\phi_3), 2)] \quad (57)$$

and $W_3(\phi_3) = y(\phi_3, 3)$. To understand $W_2(\phi_3)$, note that after success in job 2, the prior is $P_{H2}(\phi_3) \geq \rho_3$, so working at job 3 is optimal. After a failure in job 2, the prior is $P_{L2}(\phi_3)$ and the second no demotion condition in (54) implies that job 2 is optimal in the second period. The reason is that if instead the worker were assigned to job 3, then there would be no value to information at ϕ_3 , so $\phi_3 = \rho_3$, which contradicts the no demotion condition.

Now consider the implications for wages: the successful workers receive wage increases and the unsuccessful workers receive wage decreases. Thus, this example, though simple, implies many of the patterns observed in the BGH data. Promotions are by only one step, wages increase upon successful performance and promotion, no demotion occurs, but wage decreases are possible.

Finally, the promotion and demotion rungs can be described. The demotion rungs are trivial: no workers are demoted. There are two promotion rungs for workers assigned to job 1, namely,

$$[0, \phi_2) = [0, q_H(\rho_2, \theta_1)) \cup [q_H(\rho_2, \theta_1), \phi_2),$$

where $q = q_H(\rho_2, \theta_1)$ is the prior such that a success in job 1 leads the prior to be updated to the static cutoff for job 2, that is, q solves $P_H(q, \theta_1) = \rho_2$. All promoted workers, those with priors in $[q_H(\rho_2, \theta_1), \phi_2)$, receive wage increases. Notice that a worker with prior $p \in [\rho_2, \phi_2)$ is assigned to job 1 and paid $y(p, 1)$, which is below the worker's current expected output in job 2. Hence, when workers with such priors are promoted, their wages experience an extra jump in wages. Likewise, there are two promotion rungs for workers assigned to job 2, namely,

$$[\phi_2, \phi_3) = [\phi_2, q_H(\rho_3, \theta_2)) \cup [q_H(\rho_3, \theta_2), \phi_3),$$

where $q = q_H(\rho_3, \theta_2)$ is the prior such that a success in job 2 leads the prior to be updated

to the static cutoff for job 3, that is, q solves $P_H(q, \theta_2) = \rho_3$. For $p \in [\phi_2, q_H(\rho_3, \theta_2))$, after a success in job 2 a worker stays in job 2 and for $p \in [q_H(\rho_3, \theta_2), \phi_3)$, after a success in job 2 a worker is promoted to job 3. Finally, when workers with $p \in [\rho_3, \phi_3)$ are promoted, their wages experience an extra jump.

6 Human Capital Acquisition

The existing framework is promising in accounting for many features of observed job assignments and wages. One dimension of the data the model may not capture is that individual wages on average increase with time in the labor market. In the current framework, two workers with the same prior have the same wage regardless of their experience, measured by the number of periods both have been in the labor market. This feature can be easily remedied by allowing for human capital acquisition. I argue that adding human capital accumulation leads to only minor modifications in the existing analysis and can bring the model more in line with the data.

The simplest way to do so is to suppose it depends only on experience t . Suppose first that experience just adds a positive constant to output at any job, so that now expected output is given by

$$y_f(p, k, t) = y_f(p, k) + \gamma t \quad (58)$$

with $\gamma \geq 0$. Then, it should be clear that the earlier assignment rules are unchanged and that the earlier expressions for wages are simply augmented by γt . That is, human capital acquisition adds a drift term to wages and has no other effect. Specifically, let the *human capital* economy be the earlier economy with $y_f(p, k, t)$ in (58) replacing $y_f(p, k)$. It is easiest to state the characterization result for the general setup.

Proposition 7. *The optimal assignment rule in the human capital economy solves (16) and hence coincides with that of the general economy. Equilibrium wages in the human capital economy are given by*

$$w(p_t, \theta_{fk}, t) = w(p_t, \theta_{fk}) + \gamma t, \quad (59)$$

where $w(p_t, \theta_{fk})$ is given by (15).

Proof: The proof that the assignment rules for the two economies coincide follows immediately from the observation that the objective function for the planning problem in the human capital economy is the sum of the objective function for that problem in the original economy plus terms that are independent of the allocations of workers to jobs and firms. Hence, the two solutions coincide. The proof that wages are given by (59) follows from the observation that the zero profit condition implies the analog of (14). Thus, (59) follows using the same logic that was used to derive (15). \square

Here, I have considered the simplest case in which human capital is produced and used in the same way in all jobs. Of course, once I allow human capital to be differentially acquired in some jobs and differentially productive in other jobs, the job assignment rule in the economy with information and human capital acquisition must be dynamically designed to balance both. Hence, it will differ from the job assignment rules considered here in the obvious ways.

7 Conclusion

In this paper, I have integrated and extended two leading approaches to the analysis of careers within and across firms, so as to generate job and wage paths that resemble those observed in the data. In particular, workers commonly advance through the hierarchy of jobs of a firm and turn over between firms in pursuit of better matches of their skills to jobs and firms. By doing so, they earn higher wages but repeatedly face the possibility of failing and experiencing wage decreases.

I have characterized equilibrium assignment and wages in this integrated framework and linked the predictions of the model to existing empirical work. The results I derived apply to differing degrees of generality of ability across jobs and firms, ranging from firm specific to occupation specific to completely general across jobs and firms. Versions of the model seem promising to qualitatively explain a broad set of empirical findings about job and wage mobility in firms and in the labor market.

A Appendix

Here, I present omitted details and arguments.

A.1 An Overlapping Generations Interpretation

Consider an overlapping generations structure similar to that in Blanchard (1985). Each worker alive in period t faces a constant probability of dying of $1 - \gamma$ with $\gamma \in (0, 1)$. At time t the probability of being alive at $t + s$ is thus γ^s . In each period, a measure of agents are born with a distribution of priors $G(p_1)$. The measure of such agents $\int dG(p)$ is normalized to $1 - \gamma$. By the law of large numbers, the measure of any given cohort of workers will decline deterministically with time at rate $1 - \gamma$. Using $m_{t,s}$ to denote the measure in period s of a cohort born in period t , we have $m_{t,s} = (1 - \gamma)\gamma^{s-t}$. Thus, the measure of the total population of workers at time s is $m_s = \sum_{t=-\infty}^s m_{t,s}$, which is constant and equal to 1.

Following Blanchard (1985), I also assume the existence of perfect annuity markets. Firms in these markets make annuity payments to workers when they are alive and inherit the wealth of these workers when they die. I also assume the existence of complete markets for insurance

against the idiosyncratic risk of success or failure in a job. Since there is no aggregate uncertainty in this economy, the combination of annuity markets and insurance markets renders the problem of a consumer deterministic.

At birth, workers differ only in their initial prior p_1 . Hence, workers' consumption paths will be indexed by their initial priors. Let $\tilde{\delta}$ denote a worker's discount factor ignoring the possibility of death. Then, the workers's effective discount factor is $\delta = \tilde{\delta}\gamma$. A worker born at time s with initial prior p_1 has utility given by

$$(1 - \delta) \sum_{t=s}^{\infty} \delta^{t-s} u(c_{s,t}(p_1)),$$

where $c_{s,t}(p_1)$ is the consumption of a worker of cohort s in period t and the period utility function $u(c_{s,t}(p_1))$ is a strictly concave function of $c_{s,t}(p_1)$. (Of course, over time the worker's prior will be updated as before, so p_1 is just an index for the initial prior and does not represent the time t prior given the worker's history of successes and failures.)

Consider a stationary equilibrium for this economy. In a stationary equilibrium, the intertemporal price of goods will be constant in calendar time. In particular, the price of goods at time t in units of goods at time s will be $1/R^{t-s}$ with $R = 1/\delta$ for the standard reasons. The budget constraint of a worker in this stationary equilibrium can be written as

$$\sum_{t=s}^{\infty} \delta^{t-s} c_{s,t} = W_s(p_1),$$

where $W_s(p_1)$ is the expected present discounted value of wages using the market interest rate R . The worker's problem conveniently divides into two parts: an *intertemporal consumption* problem and a *job choice* problem. The optimal path of consumption satisfies

$$\frac{\delta^{t-s} u'(c_{s,t})}{u'(c_{s,s})} = \frac{1}{R^{t-s}},$$

which, in this stationary equilibrium with $R = 1/\delta$, reduces to $u'(c_{s,t}(p_1)) = u'(c_{s,s}(p_1))$ so that workers smooth their consumption. The job choice problem is to choose a sequence of jobs so as to maximize the present discounted value of wages $W_s(p_1)$, where these wages are discounted at the market rate $R = 1/\delta$. That is, the job choice problem reduces to that in (11).

Hence, I have shown that the job assignment problem solved in the body of the paper also solves the job assignment problem in an overlapping generations economy with risk-averse workers and complete markets.

A.2 Omitted Proof

Proof of Corollary 2: For (i) suppose that a worker in job k at prior $p_t \in [\rho_k, \rho_{k+1})$ experiences a success, so the new prior is $P_H(p_t) \geq p_t$. If $P_H(p_t) < \rho_{k+1}$ then the worker stays in job k , since output at each job is monotone in the prior, and the wage increases from $y(p_t, k)$ to $y(P_H(p_t), k)$. If $P_H(p_t) \geq \rho_{k+1}$, the worker is promoted to job $k+1$ and the new wage is $y(P_H(p_t), k+1)$. That this wage is greater than $y(p_t, k)$ follows by monotonicity, which implies $y(P_H(p_t), k) \geq y(p_t, k)$,

and by the single-crossing property, which implies that $y(p_t, k + 1) > y(p_t, k)$ if $p_t > \rho_k$. The proof of (ii) is symmetric. Result (iii) immediately follows from monotonicity and single-crossing. For (iv), I formalize the statement that conditional on current wages, wage increases are positively serially dependent as

$$\Pr(w_{t+2} - w_{t+1} > 0 | w_{t+1} - w_t > 0, w_t) \geq \Pr(w_{t+2} - w_{t+1} > 0 | w_t). \quad (60)$$

By the fact that wages increase from any period t to $t + 1$ only if a success occurs, an event denoted by $z_t = H$, and that by (29) wages are increasing in the prior, (60) can be rewritten as

$$\Pr(z_{t+1} = H | z_t = H, p_t) \geq \Pr(z_{t+1} = H | p_t). \quad (61)$$

Since the left side of (61) equals $\Pr(z_{t+1} = H | P_H(p_t))$, the result follows since the probability of success increases in the prior. Next, note that (v) means that, given $k_{t+2} = k'' > k' > k_t = k$, and the fact that the worker is assigned to job k at t and receives the wage w_t , the following inequality holds

$$\Pr(k_{t+2} = k'' > k_{t+1} | k_{t+1} = k' > k_t, w_t) \geq \Pr(k_{t+2} = k'' > k_{t+1} | k_{t+1} = k, w_t). \quad (62)$$

Here, I restrict attention to wages w_t such that the probabilities in (62) are well defined, that is, wages at job k_t in t associated with priors such that the probability of promotion to job $k_{t+1} > k_t$ is positive. Let p_t denote a prior consistent with w_t and job k_t in this sense. The conditioning statement on the left side of (62), that a promotion occurred at $t + 1$ starting from some prior p_t in t , implies that $p_{t+1} = P_H(p_t)$ by (1). The event that $k_{t+2} = k'' > k_{t+1} = k'$ requires that a success also occurred in $t + 1$ at job k_{t+1} so that $p_{t+2} = P_H(p_{t+1})$ with $P_H(p_{t+1}) \geq \rho_{k_{t+2}} > p_{t+1} \geq \rho_{k_{t+1}}$. Then, either both sides of (62) are zero or two successes from p_t lead the prior to be $p_{t+2} = P_H(P_H(p_t)) \geq p_{k_{t+2}}$, in which case the result clearly holds, since the statement can be rewritten $\Pr(z_{t+1} = H | P_H(p_t)) = \Pr(z_{t+1} = H | z_t = H, p_t) \geq \Pr(z_{t+1} = H | p_t)$.

As for (vi), note that for a promotion to occur from k to $k + 1$ in period $t + 1$, p_t must be such that $P_H(p_t) \geq \rho_{k+1} > p_t$. As β approaches α , the difference $P_H(p_t) - p_t$ becomes arbitrarily small. Hence, promotions from k to $k + 1$ occur only at p_t just below ρ_{k+1} leading to $P_H(p_t)$ just above ρ_{k+1} . Since wages equal expected output by (14), by the definition of ρ_{k+1} it follows that as β approaches α , wage changes become arbitrarily small upon promotion. At level k the wages range from $w(\rho_k, k) = y(\rho_k, k)$ to (just below) $w(\rho_{k+1}, k + 1) = y(\rho_{k+1}, k + 1)$. Since it is possible to adjust y_{Hk} and y_{Lk} to keep this range fixed while β is made closer to α , the result follows. \square

References

- Anderson, A., and L. Smith** (2010): “Dynamic Matching and Evolving Reputations,” *Review of Economic Studies* 77(1), 3–29.
- Baker, G., M. Gibbs, and B. Holmström** (1994a): “The Internal Economics of the Firm: Evidence from Personnel Data,” *Quarterly Journal of Economics* 109(4), 881–919.
- (1994b): “The Wage Policy of a Firm,” *Quarterly Journal of Economics* 109(4), 921–955.
- Banks, J.S., and R.K. Sundaram** (1992): “Denumerable-Armed Bandits,” *Econometrica*

60(5), 1071–1096.

Bernhardt, D. (1995): “Strategic Promotion and Compensation,” *Review of Economic Studies* 62(2), 315–339.

Blanchard, O. (1985): “Debt, Deficits, and Finite Horizons,” *Journal of Political Economy* 93(2), 223–247.

Buchinsky, M., D. Fougère, F. Kramarz, and R. Tchernis (2010): “Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the United States,” *Review of Economic Studies* 77(3), 972–1001.

Eeckhout, J., and W. Xi (2009): “Assortative Learning,” mimeo.

Felli, L., and C. Harris (1996): “Learning, Wage Dynamics, and Firm-Specific Human Capital,” *Journal of Political Economy* 104(4), 838–868.

Felli L., and C. Harris (2004): “Firm-Specific Training,” STICERD Theoretical Economics, Discussion Paper No. TE/04/473.

Ferrall, C. (1997): “Empirical Analysis of Occupational Hierarchies,” *Journal of Human Resources* 32(1), 1–34.

Ferrall, C., K.G. Salvanes, and E.Ø. Sørensen (2009): “Wages and Seniority When Coworkers Matter: Estimating a Joint Production Economy Using Norwegian Administrative Data,” QED Working Paper 1200.

Flinn, C.J. (1986): “Wages and Job Mobility of Young Workers,” *Journal of Political Economy* 94(3), 88–110.

Gibbons, R., and M. Waldman (1999): “A Theory of Wage and Promotion Dynamics Inside Firms,” *Quarterly Journal of Economics* 114(4), 1321–1358.

— (2006): “Enriching a Theory of Wage and Promotion Dynamics Inside Firms,” *Journal of Labor Economics* 24(1), 59–107.

Groes, F., P. Kircher, and I. Manovskii (2010): “The U-Shapes of Occupational Mobility,” mimeo.

Holmström, B., and J. Tirole (1989): “The Theory of the Firm,” in *Handbook of Industrial Organization*, vol. 1, ed. R. Schmalensee and R. Willig, 61–133, Amsterdam: North-Holland.

Jovanovic, B. (1979): “Job Matching and the Theory of Turnover,” *Journal of Political Economy* 87(5), 972–990.

Jovanovic, B., and Y. Nyarko (1997): “Stepping-Stone Mobility,” *Carnegie-Rochester Conference Series on Public Policy* 46(1), 289–325.

Kambourov, G, and, I. Manovskii (2009a): “Occupational Mobility and Wage Inequality,” *Review of Economic Studies* 76(2), 731–759.

Kambourov, G, and, I. Manovskii (2009b): “Occupational Specificity of Human Capital,” *International Economic Review* 50(1), 63–115.

Kremer, M. (1993): “The O-ring Theory of Economic Development,” *Quarterly Journal of Economics* 108(3), 551–575.

Kremer, M., and E. Maskin (1996): “Wage Inequality and Segregation by Skill,” NBER Working Paper 5718.

MacDonald, G. (1982): “A Market Equilibrium Theory of Job Assignment and Sequential Accumulation of Information,” *American Economic Review* 72(5), 1038–1055.

- Miller, R.A.** (1984): “Job Matching and Occupational Choice,” *Journal of Political Economy* 92(6), 1086–1120.
- Moscarini, G., and F. Vella** (2008): “Occupational Mobility and the Business Cycle,” mimeo.
- Neal, D.** (1995): “Industry-Specific Human Capital: Evidence from Displaced Workers,” *Journal of Labor Economics* 13(4), 653–677.
- Neal, D.** (1999): “The Complexity of Job Mobility Among Young Men,” *Journal of Labor Economics* 17(2), 237–261.
- Parent, D.** (2000): “Industry-Specific Capital and the Wage Profile: Evidence from the National Longitudinal Survey of Youth and the Panel Study of Income Dynamics,” *Journal of Labor Economics* 18(2), 306–323.
- Pastorino, E.** (2013): “Careers in Firms: Estimating a Model of Job Assignment, Learning, and Human Capital Acquisition,” mimeo.
- Prescott, E. C., and M. Visscher** (1980): “Organization Capital,” *Journal of Political Economy* 88(3), 446–461.
- Rosen, S.** (1982): “Authority, Control, and the Distribution of Earnings,” *Bell Journal of Economics* 13(2), 311–323.
- Rubinstein, Y., and Y. Weiss** (2007): “Post Schooling Wage Growth: Investment, Search and Learning,” in *Handbook of the Economics of Education*, vol. 1, ed. E.A. Hanushek and F. Welch, 1–67, Amsterdam: North-Holland.
- Topel, R.** (1991): “Specific Capital, Mobility, and Wages: Wages Raise with Job Seniority,” *Journal of Political Economy* 99(1), 145–176.
- Waldman, M.** (1984): “Job Assignments, Signalling, and Efficiency,” *RAND Journal of Economics* 15(2), 255–267.

Figure 1. Expected Output and Benchmark Assignment Rule

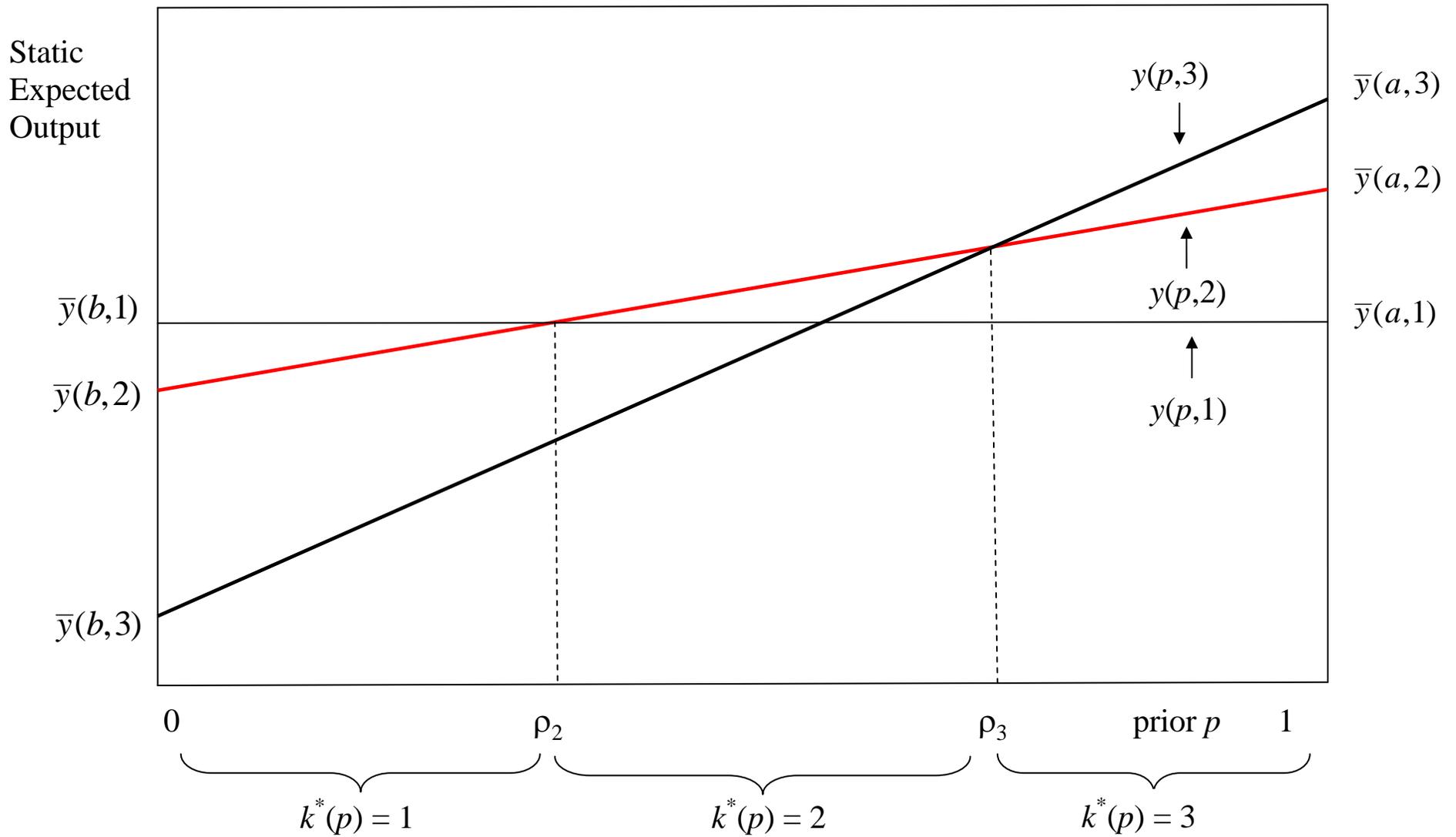
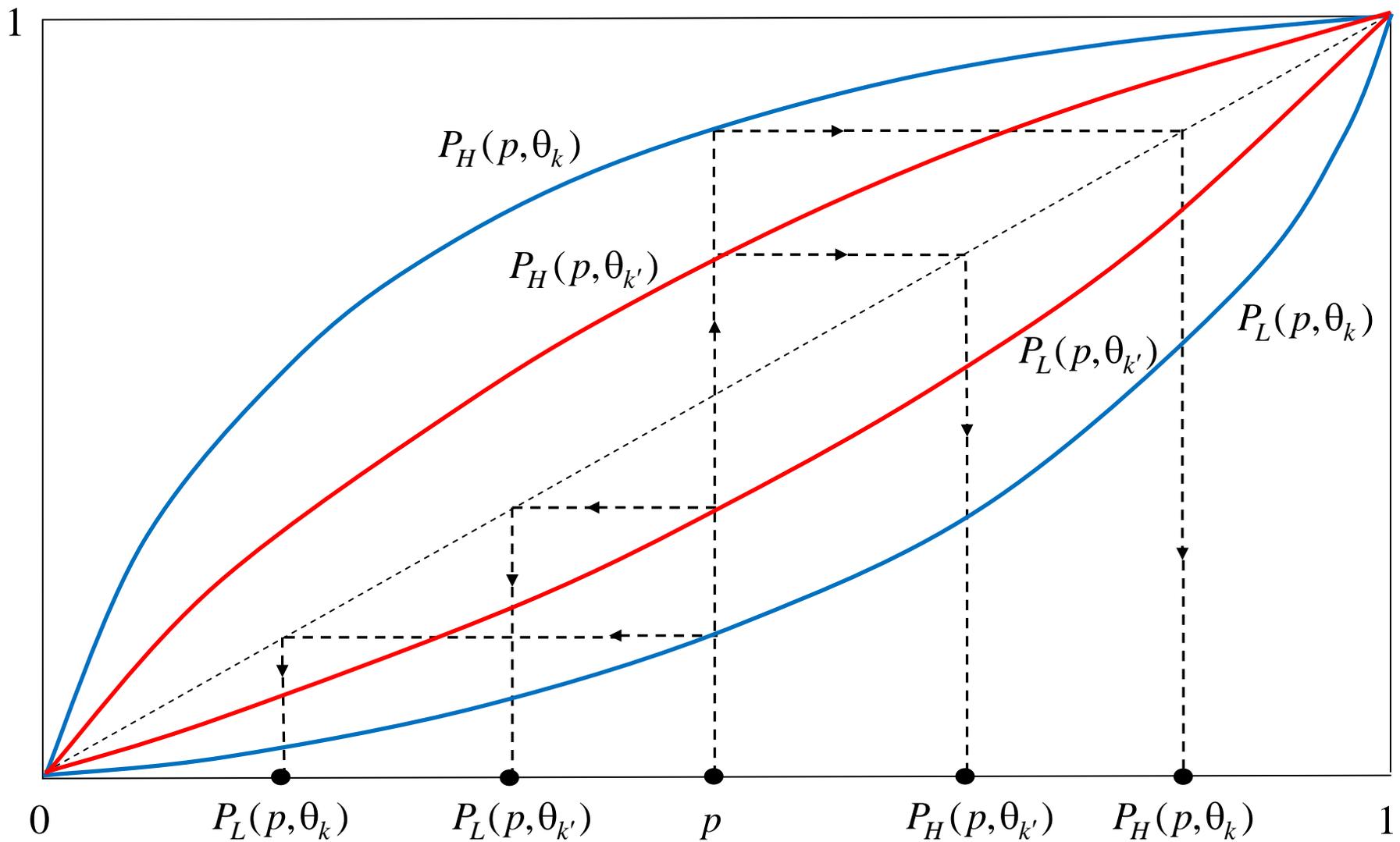
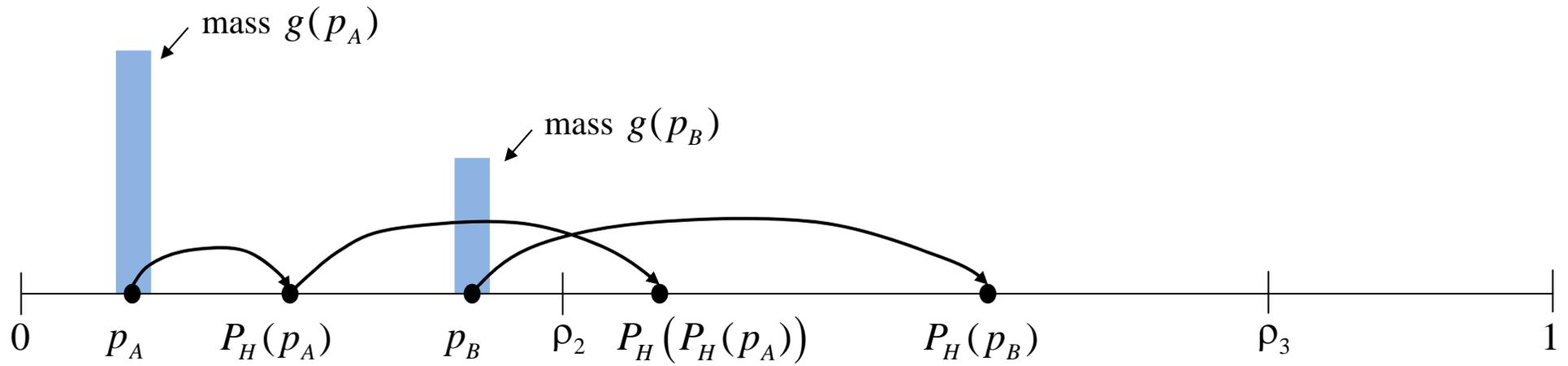


Figure 2. Update when Job k is More Informative than Job k'



Note: The posteriors $P_L(p, \theta_k)$ and $P_H(p, \theta_k)$ under the more informative job are a mean-preserving spread of the posteriors $P_L(p, \theta_{k'})$ and $P_H(p, \theta_{k'})$ under the less informative job.

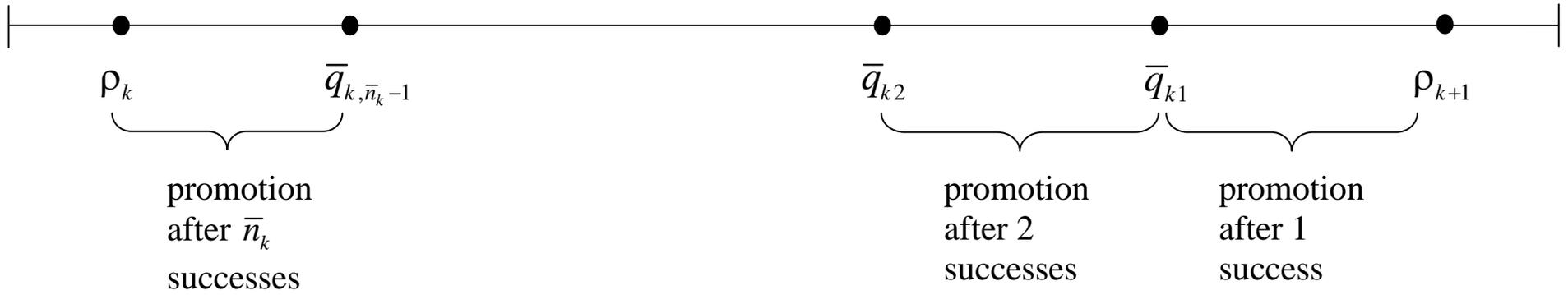
Figure 3. Heterogeneous Priors and Promotions



Note: Measure $g(p_A)$ of workers have prior p_A and measure $g(p_B)$ of workers have prior p_B .

Figure 4. Promotion and Demotion Rungs

A. Promotion Rungs



B. Demotion Rungs



Figure 5. Multi-Step Promotions

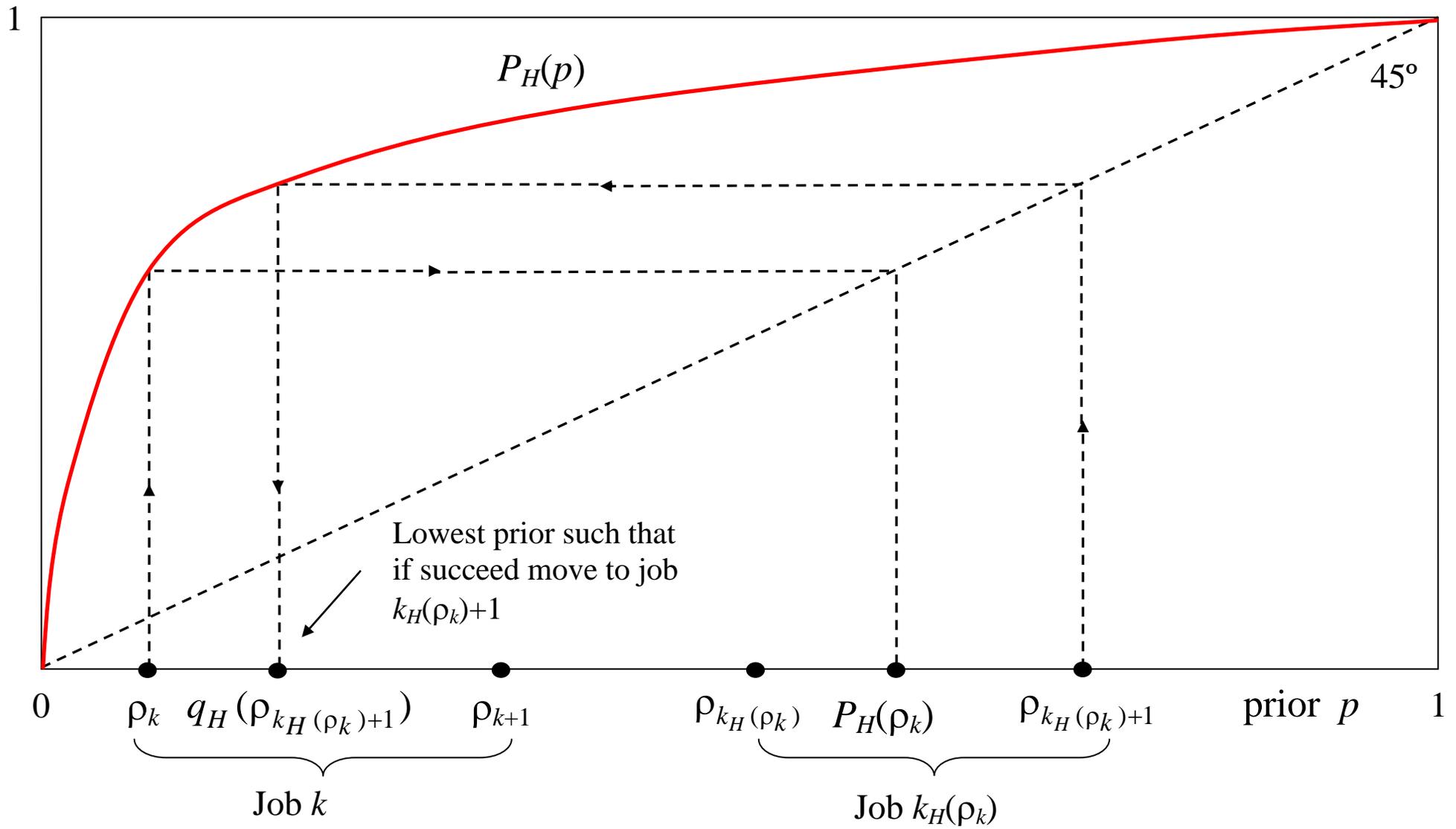


Figure 6. Information Benefits of Jobs: Job k More Informative Than Job k'

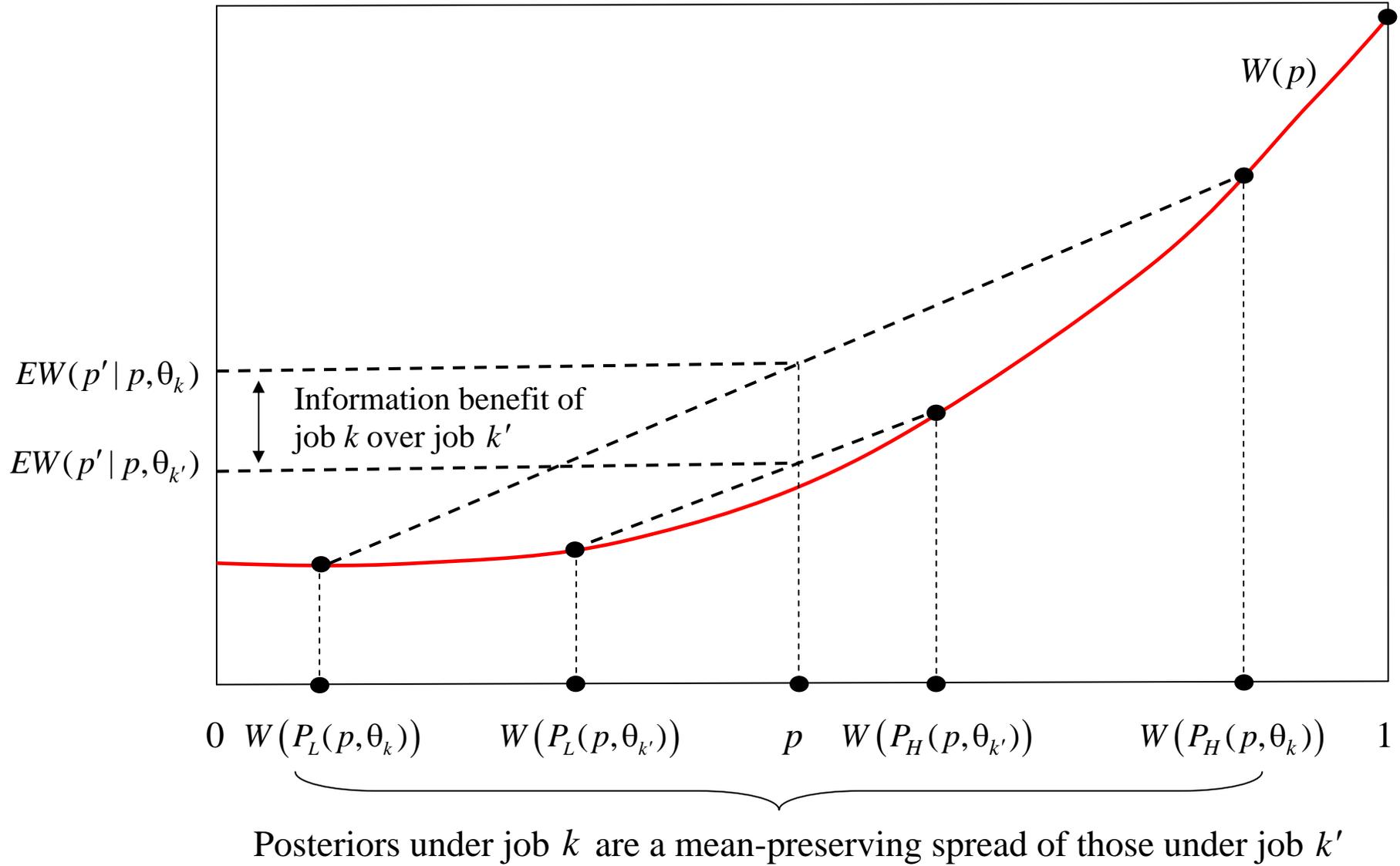
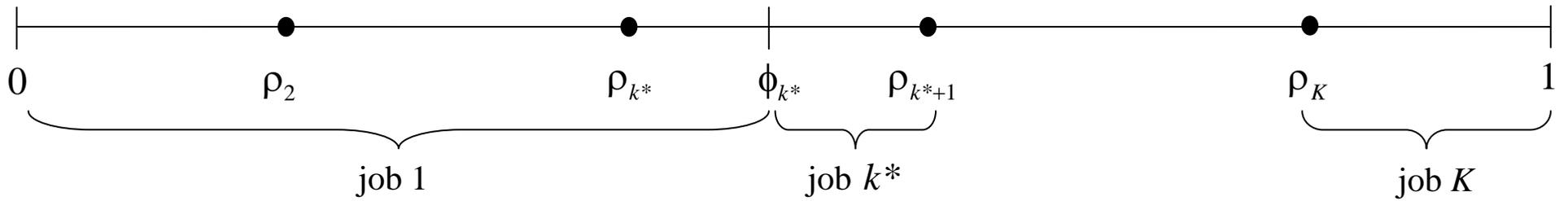


Figure 7. Assignments in the Entry Job Economy

A. $\phi_{k^*} < \rho_K$



B. $\phi_{k^*} \geq \rho_K$

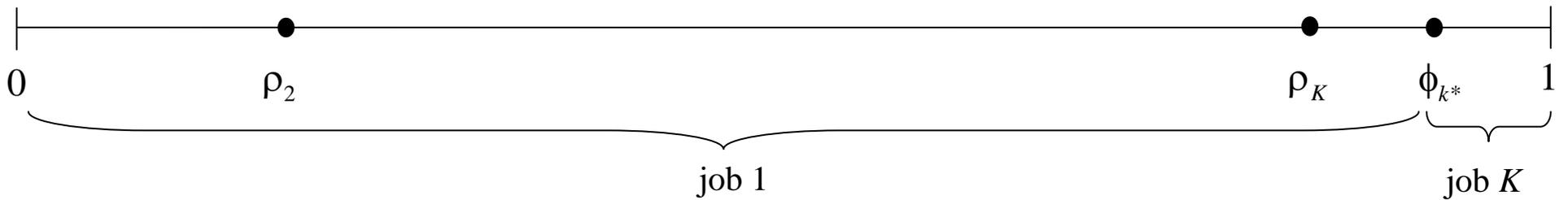


Figure 8. The Entry Job Economy

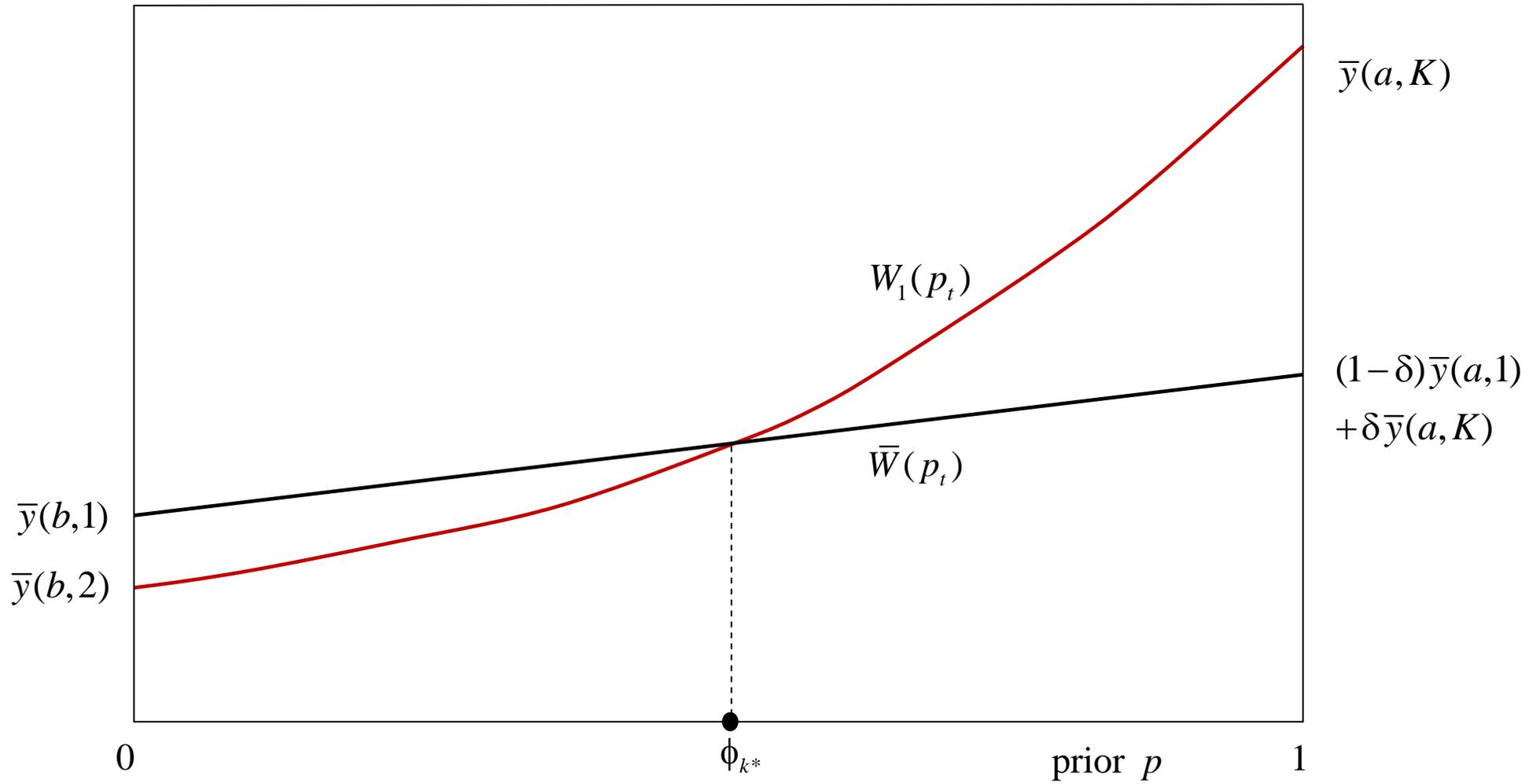
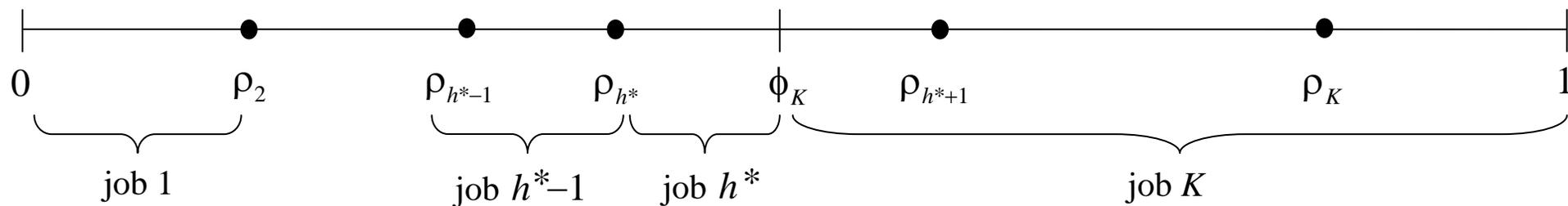


Figure 9. Assignments in the Top Job Economy

A. $\phi_K \geq \rho_2$



B. $\phi_K < \rho_2$



Figure 10. A Two-Period Example

