Federal Reserve Bank of Minneapolis
Research Department Staff Report 469

July 2012

Careers in Firms: Estimating a Model of Learning, Job Assignment, and Human Capital Acquisition*

Elena Pastorino

Federal Reserve Bank of Minneapolis
and University of Minnesota

ABSTRACT

This paper develops and structurally estimates an equilibrium model of the labor market that integrates learning, job assignment, and human capital acquisition to account for the main patterns of job and wage mobility characteristic of careers in firms. A key innovation is the modeling of firms' incentives to experiment that arise from the ability of firms, through job assignment, to affect the rate at which they acquire information about workers. The resulting trade-off between output and information implies that a firm's retention and job assignment policy solves an experimentation problem: a so-called multi-armed bandit with dependent arms. The model is estimated using longitudinal administrative data from one U.S. firm in a service industry (the same data used by Baker, Gibbs, and Holmström (1994a,b)) and fits the data remarkably well. My estimates indicate that learning during employment accounts for a significant fraction of measured wage growth on the job, whereas experimentation through job assignment primarily contributes to explaining the patterns of job mobility within the firm. Since learning is gradual, however, persistent uncertainty about workers' abilities is responsible for a substantial compression of wage growth with tenure.

Keywords: Experimentation; Bandit; Wage Growth; Job Mobility; Turnover
JEL Classification: D22, D83, J24, J31, J44, J62

*I am truly indebted to Ken Wolpin for his generous advice and encouragement. I benefited from numerous conversations with Petra Todd and Antonio Merlo. I am very grateful to John Geweke and Patrick Kehoe for their suggestions and careful reading of the manuscript. Andrew Ching, Hiro Kasahara, Matt Shum, and Michael Waldman have provided me with valuable comments on the current version of the manuscript. This version also incorporates many useful suggestions by the co-editor and three anonymous referees. Finally, I thank George Baker for kindly providing me with the data and Bengt Holmström for his support in the project. Zhen Huo provided outstanding research assistance, and Joan Gieseke and Kathy Rolfe 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 the past three decades, a large body of work has attempted to understand the organization of production in firms and its implications for *careers in firms*, that is, the allocation of workers to tasks and jobs within a firm and the dynamics of wages with tenure. Based on the empirical implications of existing models and accumulating descriptive evidence on the characteristics of careers in firms, a consensus has emerged that a theoretical framework that combines learning, job assignment, and human capital acquisition is a promising candidate for explaining a broad set of empirical findings related to job and wage mobility in firms. (See Gibbons and Waldman (1999a,b) and Waldman (forthcoming).) So far, however, no comprehensive examination of the empirical relevance of such a model or of the relative importance of its components has been systematically conducted.

This integrated framework for the analysis of careers in firms is also known to be missing two crucial elements: experimentation within firms and worker turnover between firms. Specifically, the existing framework models learning about workers' abilities as a process of passive information acquisition, based on productivity signals generated through employment. As such, it ignores the possibility that firms may actively acquire information by choosing, through job assignment, the precision of the information they gather about employed workers; that is, firms may *experiment*. The existing framework also largely abstracts from the fact that workers typically experience employment at different firms. Both of these elements, experimentation and turnover, are central to understanding careers in firms and the connection between the dynamics of jobs and wages with tenure at a firm and with experience in the labor market. To date, this connection has neither been well understood theoretically nor much investigated empirically.

In this paper, I develop and structurally estimate an equilibrium model of the labor market that extends existing models of learning, job assignment, and human capital acquisition to incorporate experimentation within firms and turnover between firms, in order to account for salient features of individual careers in firms. Based on the estimation of this integrated model with data on the careers of managers from a single U.S. firm in a service industry (the same data used by Baker, Gibbs, and Holmström (1994a,b)), I show that this model well accounts for the main patterns of job and wage mobility in firms. The estimated model also implies that learning and experimentation are quantitatively important sources of observed career paths. Finally, I find that persistent uncertainty about ability, resulting from gradual learning, compresses average wages and wage growth with tenure.

The idea that experimentation plays an important role in careers in firms dates back to Prescott and Visscher (1980) and Holmström and Tirole (1989). These authors argue that different jobs within firms typically have different information content in that performance at different jobs may provide different amounts of information about a worker's ability. Thus, by choosing a job for a worker, a firm also chooses the amount of information it acquires about the worker's ability. By allowing for such experimentation, I account for a dimension that Holmström and Tirole (1989) view as critical but missing from existing analyses of labor markets internal to firms: in practice, firms assign workers to jobs with such differential learning possibilities in mind. Moreover, Baker, Gibbs, and Holmström (1994a), henceforth *BGH*, provide an empirical motivation for incorporating this feature. Based on their data, they argue that experimentation through job assignment is likely to be a major source of observed job and wage dynamics in firms.

The idea that turnover is a pervasive feature of individual careers is well known and has been amply documented. For instance, in the BGH data I use, over 60 percent of the managers left the firm BGH study after just eight years. However, existing analyses of careers in firms typically abstract from it, thus obscuring the factors that affect separations between firms and workers and, in turn, wage growth with tenure.[1] In contrast, by including turnover in the model, I can provide a mechanism for the link between the sources of job and wage mobility within firms and those between firms that can naturally account for the joint behavior of workers' wages, tenure, and turnover.

Formally, my model incorporates experimentation and turnover within an equilibrium model of the labor market that features learning, job assignment, and human capital acquisition with firm tenure and labor market experience. My analysis of firm behavior is based on a specification of production in firms as organized in distinct *jobs* and starts with the job assignment problem that a firm faces when employing a worker. Initially, an individual worker's ability is unobserved by that worker as well as by all firms, but they jointly learn about this ability over time by observing the worker's performance in the assigned jobs. I assume that jobs differ in the impact of a worker's ability on output and in the precision of the information that performance conveys about ability. Crucially, I model a worker's ability as correlated across jobs so that successful performance in a job may lead a firm to optimally assign the worker to a different job at which ability is more valuable. With employment at a given job in a firm, a worker also acquires new productive skills, or *technological human capital*, that can be to a varying degree *task-specific* in that these skills are only partly transferable to other jobs within the firm or *firm-specific* in that they are only partly transferable to jobs in other firms.

Consider now how the model gives rise to experimentation. Since jobs differ in their output and in how informative performance is regarding ability, a firm determines a worker's job assignment by trading off the worker's current expected output in a job against the value of the information conveyed by performance and of the additional human capital the worker accumulates with experience in that job. Similarly, in deciding which job and wage offer to

[1] An exception is Ghosh (2007), whose model of learning, human capital acquisition, and competition among homogeneous firms produces turnover in equilibrium due to random disutility shocks to continuing employment with a same firm.

accept, for each offer a worker trades off the current wage with the benefit of the information and human capital that can be acquired at the proposed job. The intricate dynamic problem that each firm and worker solve, due to this trade-off between output and information, is a classic *experimentation problem*, a so-called multi-armed bandit with *dependent arms*. (See Easley and Kiefer (1988) for an illustration.) Specifically, differently from the standard case of independent alternatives, that is, arms, here the correlation of a worker's ability across jobs implies that the arms of the bandit are statistically dependent.

Consider next how my model gives rise to turnover. First, the model incorporates exogenous turnover through separation shocks. Second, it generates endogenous turnover through the interplay of the heterogeneity in firm technologies, the updating of priors about workers' abilities, and Bertrand competition among firms for workers. More precisely, different technologies across firms imply that the expected present discounted value of output of a worker with a given prior differs across firms. As performance signals accumulate, the prior about the worker's ability is updated and the firm for which that worker is most profitable may then change. Because of competition, the firm for which the worker is currently most profitable can successfully bid the worker away from the worker's previous employer, if this latter firm, at the new prior, is no longer the one for which the worker is most profitable. Hence, workers naturally reallocate across firms as their priors evolve.

In the paper I characterize equilibrium in a way that is amenable to the estimation of the model based on my data on a single firm. First, I show that in equilibrium each firm's best-response problem can be conveniently combined with a worker's best-response problem into a joint dynamic assignment problem. I refer to this problem as the *match surplus* maximization problem between a firm and a worker. In equilibrium, the solution to this problem determines both the states at which a firm employs a worker and, if so, the job to which the worker is assigned. Hence, equilibrium allocations are solutions to the collection of match surplus maximization problems between each firm and worker.

Second, I rely on equilibrium to derive a simple and intuitive expression for paid wages. The wage paid by an employing firm must compensate a worker for the loss of the information and human capital that the worker could have acquired by employment at other firms. Thus, the paid wage not only reflects the worker's expected output at the best competitor of the employing firm, as in a standard Bertrand game of wage competition, but it also includes a compensating premium for the worker's forgone prospects of learning and human capital acquisition.

Together, the match surplus problem and the expression for the paid wage for each worker fully characterize equilibrium from the point of view of any given firm. As such, they provide a convenient basis for estimating the model on the BGH data in a way that fully accounts for the endogeneity of employment, job assignment, and wage decisions by firms and workers. Indeed, by relying on the equilibrium logic disciplining the patterns of turnover, job assignments, and wages, I can naturally correct for potential issues of selection when measuring within-firm wage growth.[2] Modeling equilibrium also allows me to assess the degree of monopsony power of the firm in my data.

As for the empirical analysis, in estimation I augment the model to allow for additional dimensions of unobserved heterogeneity among workers. By so doing, in accounting for observed job and wage mobility, I can flexibly distinguish between the role of incomplete information about ability and the role of self-selection by firms and workers based on attributes and outcomes unobserved to the econometrician. I estimate this augmented model by full-solution, full-information, nonparametric maximum likelihood using longitudinal administrative information about the careers of all managers of a single U.S. firm in a service industry, the data of BGH. Crucially for the estimation of the learning component of the model, the data contain information not only on each manager's job level and wage in each year of employment but also on the outcome of the yearly evaluation of a manager's performance. I estimate the model's parameters using eight years of observations on the cohorts of managers entering the firm at the lowest managerial level between 1970 and 1979.

As is well known, estimating a learning model presents special challenges relative to complete information models. These challenges arise from the rich data required to uncover informational structures, the presence of dynamic selection based on unobserved endogenous state variables, and the computational difficulty of incorporating serially correlated disturbances. (See the reviews of Miller (1997) and Ching, Erdem, and Keane (2011).) My exercise faces two additional challenges: the estimation of an equilibrium model in which both sides of the market are fully forward-looking, strategic, and face nonstandard experimentation problems, and the focus of estimation on multiple observed outcomes of interest, namely, job assignments, performance evaluations, and wages, which are jointly determined by several time-dependent dimensions of heterogeneity unobserved to the econometrician.

Notwithstanding these challenges, I show that, based on my equilibrium characterization, standard dynamic discrete choice techniques for single-agent decision problems can be applied to the match surplus problem between each firm and worker. Hence, in my setup, information about individual careers at one firm is sufficient to recover explicit empirical measures for firm and worker learning that are theoretically motivated. In turn, these measures can be used to counterfactually assess the importance of learning, experimentation, and persistent uncertainty about ability for job and wage profiles.

In terms of results, despite its parsimony and tight theoretical restrictions, the estimated

[2] In a similar spirit, Buchinsky, Fougère, Kramarz, and Tchernis (2010) underscore the importance of jointly estimating the determinants of employment, mobility decisions, and wages, in order to correctly measure wage growth on the job.

model successfully captures the multiple nonlinear, nonmonotone patterns of transitions of managers across the levels of the firm's job hierarchy in my data. The model also well replicates the distribution of performance ratings at the main levels in each tenure year. Lastly, the model closely matches the distribution of wages at each level of the job hierarchy and their evolution with tenure at the firm.

The estimates of the model's parameters imply several key characteristics of the process of information acquisition at the firm. First, initial uncertainty at the time of a manager's entry into the firm proves to be substantial: over half the managers in my data have initial priors about ability being high (as opposed to low) close to 0.50. Second, initial priors about ability are highly heterogeneous across managers, implying a significant dispersion in information at hiring. Third, learning is estimated to be a very gradual process: more than 15 years of good performance are necessary for the average prior about a manager's ability being high to converge to 0.90. Together, these three features imply that the degree of uncertainty about managers' ability is large, diverse across managers, and highly persistent over time.

My estimates also have implications for the process of human capital acquisition at the firm. First, I document a substantial cost to demoting a manager from a higher level to a lower level job. Second, I estimate a hump-shaped pattern of human capital accumulation with tenure at the firm, which largely accounts for the hump-shaped pattern of the hazard rate of promotion to higher levels with tenure. Third, I find that the transferability of acquired human capital across levels within the firm differs by level. In particular, human capital acquired at the highest level is much more task-specific than that acquired at lower levels.

Finally, by recovering prior beliefs and parameters of the distribution of performance at each level, I can measure the importance of learning, experimentation, and persistent uncertainty about ability for the joint dynamics of job assignments and wages. Three main findings emerge. The first finding is that learning contributes significantly to wage growth on the job: it accounts for 26 percent of cumulative wage growth over the first seven years of tenure at the firm. The second finding is that when jobs are equally informative, so that experimentation is precluded, managers are much more quickly promoted to high-level jobs even though learning is estimated to occur slowly over time. An interesting implication of this finding is that the observed pace of promotions reflects primarily the relative informativeness of different jobs and, hence, the scope for experimentation rather than the absolute speed of learning in jobs. The third finding, obtained by comparing the estimated wage growth with the wage growth arising under alternative scenarios of faster learning, is that persistent uncertainty about ability is responsible for a substantial compression of wage levels and wage growth with tenure. Persistent uncertainty about ability has, thus, a significantly adverse impact on managers' returns to employment.

Related Literature. Consider how my work relates to the existing literature. By integrating learning, job assignment, and human capital acquisition, my model builds on two influential papers on careers in firms: the work of Gibbons and Waldman (1999b, 2006), which I refer to jointly as *GW*. Differently from GW, my model incorporates both experimentation and turnover. GW abstract from experimentation by assuming that all jobs are equally informative about ability, so that a firm's job assignment problem reduces to a static problem. GW also abstract from turnover by endowing all firms with identical production technologies (and assuming an infinitesimal moving cost for workers).

More broadly, my model shares three key features of existing models. First, regarding job assignment, I follow Rosen (1982), Waldman (1984), and GW in allowing for the jobs of a firm's hierarchy to be ranked by individual ability, which I model as common across jobs, so that it may be profitable for a firm to assign higher ability workers to higher level jobs. Second, following Jovanovic (1979), Prescott and Visscher (1980), MacDonald (1982), Miller (1984), and GW, I incorporate in the 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*, that, contrary to Jovanovic (1979) and Miller (1984), are general across jobs and firms.[3] This assumption of symmetric observability is common in the learning literature. (See, for example, Jovanovic (1979), Prescott and Visscher (1980), Harris and Holmström (1982), Miller (1984), Bergemann and Välimäki (1996), Farber and Gibbons (1996), and GW.) Third, I integrate technological human capital in the model by assuming that workers can, over time, improve their ability to produce output. As do GW, I allow for the human capital acquired at a given job in a firm to be task-specific and firm-specific, as mentioned.

My paper is also closely related to empirical models of learning. For work in the labor literature, see Miller (1984), Flinn (1986), Berkovec and Stern (1991), and Nagypál (2007). For work in the personnel economics literature, see Lluis (2005) and Hunnes (forthcoming). Their work builds on Gibbons, Katz, Lemieux, and Parent (2005), who, based on a model similar to Gibbons and Waldman (1999b), assess the importance of comparative advantage and learning for sector-specific returns to observed and unobserved skills and for sectoral wage differentials. None of these papers, however, recover primitive parameters of learning and uncertainty. Finally, for work in the industrial organization literature, see Ackerberg (2003), Crawford and Shum (2005), and Ching (2010). Ching (2010) estimates a dynamic equilibrium learning model of the generic drug market. In his model, however, patients are myopic, experimentation is precluded, and only the quality of newly formed matches between patients and drugs is unknown.

[3]Note that Jovanovic (1979) also analyzes a multi-armed bandit problem. Since he assumes that ability is firm- (and job-) specific, his problem features *independent* rather than dependent arms. This difference is critical. Jovanovic's model predicts that firms and workers separate only after bad performance, which is counterfactual. Also, extending Jovanovic's idea to my setup by modeling matches between jobs and workers within a firm in the same way that Jovanovic models matches between jobs and workers across firms would lead to the counterfactual prediction of job changes within a firm only after unsuccessful performance.

1 The Model

In this section I describe the model, discuss the timing of events in a period, and characterize equilibrium.

1.1 The Environment

I consider a labor market in which F firms, indexed by $f = 1, 2, \ldots, F$, compete for the services of workers in each time period of an infinite horizon with periods $t \geq 1$. Since a worker enters the labor market in period 1, period t represents the number of years a worker has participated in the labor market, or the worker's *experience*. Workers inelastically supply one unit of labor each period and can be one of two types, indexed by $\theta \in \{\alpha, \beta\}$. A worker of type α is said to be of *high ability*; a worker of type β, *low ability*. These two types of workers have different ability-specific productivities, which are unobserved by all (including the given worker). For each worker, all firms and the worker share a common initial prior belief p_1 that the worker is of high ability at the beginning of period 1. Note that p_1 is a sufficient statistic for initial beliefs. Both firms and workers discount the future by the factor $\delta \in (0, 1)$.

Firms produce a homogeneous output good, which they sell in a perfectly competitive market at a price normalized to one. Each firm's technology is separable and has constant returns to scale in labor, which is the only input to production. For this reason, I can, without loss of generality, focus on the competition among firms for one worker at a time. (For explorations of the implications of complementarity in production among workers for the assignment of workers to tasks or teams, see Prescott and Visscher (1980), Kremer (1993), Kremer and Maskin (1996), and Davis (1997). For more recent work, see the comprehensive review of models and evidence on internal labor market practices by Gibbons and Waldman (1999a) and Waldman (forthcoming).)

Each firm f is endowed with a production set Y^f that consists of K^f tasks or *jobs*. (In estimation, I use observations on three jobs, so I will eventually set $K^f = 3$.) Each job k at firm f is characterized by a four-tuple $\{y_{fHk}(h_t), y_{fLk}(h_t), \alpha_{fk}, \beta_{fk}\}$ together with the productivity shock ε_{fkt}. Here, $y_{fHk}(h_t)$ and $y_{fLk}(h_t)$ represent *high output* and *low output*, the frequency of which depends on the worker's ability θ, and are thought of as the worker's *performance* in a period at firm f in job k. Note that these components of output also depend on h_t, a worker's technological human capital acquired with experience in the labor market. A high-ability worker with human capital h_t produces high output $y_{fHk}(h_t)$ in job k with probability α_{fk} whereas a low-ability worker with human capital h_t produces high output $y_{fHk}(h_t)$ in job k with probability β_{fk}. Thus, the type indices α and β are the vectors of probabilities that a high-ability and a low-ability worker produce high output: $\alpha = (\alpha_1, \ldots, \alpha_F)$ denotes the collection of F vectors $\alpha_f = (\alpha_{f1}, \ldots, \alpha_{fK^f})$, and $\beta = (\beta_1, \ldots, \beta_F)$ denotes the collection of F vectors $\beta_f = (\beta_{f1}, \ldots, \beta_{fK^f})$ with $f = 1, 2, \ldots, F$.

Since the probability of high output for a worker of either ability is allowed to vary across jobs, the informational content of performance, that is, the *informativeness of a job* as measured by the dispersion of posterior beliefs, differs across jobs. Also, since no restriction is placed on the dependence of α_{fk} and β_{fk} on the firm index, f, or the job index, k, this formulation encompasses varying degrees of specificity of ability across jobs and firms for each type of worker. At one extreme, α_{fk} varies only with the job k and not with the firm f, so the ability of a worker of type α is job-specific but general across firms. At the other extreme, α_{fk} varies only with the firm f and not with the job k, so the ability of a worker of type α is firm-specific but general across jobs. A similar logic applies to a worker of type β.

The dependence of $y_{fHk}(h_t)$ and $y_{fLk}(h_t)$ on h_t allows for the possibility that a worker may acquire human capital while employed, which affects the amount of output that the worker can produce, that is, the worker's productivity. Here, the amount of new skills that a worker accumulates through employment depends on the worker's experience in all jobs and at all firms at which the worker has been employed. More precisely, I specify acquired human capital h_t as

$$h_t = h(h_1; f_1, k_{f_1 1}; \ldots; f_{t-1}, k_{f_{t-1} t-1}) \text{ for } t \geq 2, \tag{1}$$

where $h(\cdot)$ is a function of the initial *observed* human capital h_1 of the worker at entry into the labor market as well as the history of the worker's past job assignments. When $t = 1$, $h_t = h_1$. In (1), $(f_1, k_{f_1 1}; \ldots; f_{t-1}, k_{f_{t-1} t-1})$ denotes the worker's job history from period 1 to period $t-1$, where f_t indicates the identity of the firm employing the worker in period t and $k_{f_t t}$ the assigned job at firm f_t. Note that this formulation of the human capital process allows workers of the same unobserved ability to produce different amounts of output across jobs within a firm, across firms, and over time. (In the empirical section, human capital is allowed to depend on all observable characteristics of a worker at entry into the firm I study, namely, age, education, and year of entry into the firm. Since I observe workers continuously employed at this firm, the only relevant part of the accumulating history of employment, after entry into the firm, is tenure at the firm's jobs. Empirically, only tenure at the firm and previous period job assignment have proved important. See Section 4 for details.)

The productivity shocks ε_{fkt} are additive disturbances to output that do not depend on ability. Hence, the (total) output of a worker realized at firm f, in job k, in period t is either $y_{fHk}(h_t) + \varepsilon_{fkt}$ or $y_{fLk}(h_t) + \varepsilon_{fkt}$. These productivity shocks are independent across jobs, workers, and periods. Thus, the shocks are independent across firms with different production sets. Let $\boldsymbol{\varepsilon}_{ft} = (\varepsilon_{f1t}, \ldots, \varepsilon_{fK^f t})$ denote the vector of productivity shocks at firm f in t in its K^f different jobs, and let $F(\boldsymbol{\varepsilon}_{ft})$ denote their joint cumulative distribution function. (In the empirical analysis, I assume they all have type I extreme value distributions.)

As I will elaborate, at the beginning of any period, each firm (and worker) observes all period t productivity shocks at all firms and then makes its job and wage offer. I also assume that a worker's performance is publicly observable to all. So all firms and workers share a common prior p_t that the worker is of high ability at the beginning of period t.

This assumption of symmetric observability is nearly universal in the learning literature. One interpretation of this assumption that firms (and workers) have the same information set is that all potential employers can costlessly uncover information about a worker's record of past performance, for instance, through the worker's resume and letters of recommendation. When, instead, competing firms observe a worker's job assignment but not the worker's performance, the employing firm may have an incentive to strategically delay changes in job assignment in an attempt to manipulate the market's inference about the worker's ability.[4] I think of the assumption of perfect observability of performance as a first approximation to examining the impact of competition on intra-firm job assignment and turnover in a framework in which the precision of performance information is job-dependent.

The one-period expected output of a worker with prior p_t and human capital h_t assigned to job k at firm f in t, after productivity shocks are realized, is $y_f(p_t, h_t, k) + \varepsilon_{fkt}$, where

$$y_f(p_t, h_t, k) = p_t \bar{y}_f(\alpha_{fk}, h_t, k) + (1 - p_t)\bar{y}_f(\beta_{fk}, h_t, k), \tag{2}$$

$$\bar{y}_f(\alpha_{fk}, h_t, k) = \alpha_{fk} y_{fHk}(h_t) + (1 - \alpha_{fk}) y_{fLk}(h_t), \tag{3}$$

and a similar expression holds for $\bar{y}_f(\beta_{fk}, h_t, k)$ by replacing α_{fk} with β_{fk} in (3). In (3), $\bar{y}_f(\alpha_{fk}, h_t, k)$ is the one-period expected output in job k for a worker with human capital h_t who is known to be of high ability ($p_t = 1$), and $\bar{y}_f(\beta_{fk}, h_t, k)$ is the corresponding expected output for a worker with human capital h_t who is known to be of low ability ($p_t = 0$). If $\alpha_{fk} > \beta_{fk}$ and $y_{fHk}(h_t) > y_{fLk}(h_t)$, then $y_f(p_t, h_t, k)$ is increasing in p_t.

The objective function of any firm is the expected present discounted value of profits and the objective function of any worker is the expected present discounted value of wages. I normalize at zero the outside option of a firm. Similarly, I assume that the value of the outside option of a worker is a sufficiently low constant that the worker chooses to work in the market considered in each period. For simplicity, and for reasons of model identification, I also normalize the

[4]For analyses that relax this assumption, see Greenwald (1986), Waldman (1996), and Li (2011) for models of job mobility, wage dispersion, and productivity growth and Waldman (1984, 1990), Ricart i Costa (1988), Bernhardt and Scoones (1993), and Bernhardt (1995) for models of situations in which firms other than a worker's current employer do not observe direct information about the worker's productivity but can partially infer it by considering the worker's task assignment. For other references, see the review by Waldman (forthcoming). See also DeVaro and Waldman (2012) for descriptive evidence of the potentially important signaling role of education in my data.

worker's outside option at zero.[5] Lastly, I find it convenient to normalize payoffs by $(1-\delta)$ so as to express them as per-period averages.

1.1.1 The Scope for Experimentation

Based on the performance of a worker in a job, at the end of each period all firms and workers update their priors about the worker's ability according to Bayes' rule. Specifically, if the worker's job assignment is k_{ft} at firm f in t and the output realized at the end of the period is high, then the updated prior p_{t+1} at the beginning of period $t+1$ is

$$p_{t+1} = P_{fHk}(p_t) = \frac{\alpha_{fk_{ft}} p_t}{\alpha_{fk_{ft}} p_t + \beta_{fk_{ft}}(1-p_t)}. \tag{4}$$

Similarly, if the output realized at the end of the period is low, then the updated prior p_{t+1} at the beginning of period $t+1$ is

$$p_{t+1} = P_{fLk}(p_t) = \frac{(1-\alpha_{fk_{ft}}) p_t}{(1-\alpha_{fk_{ft}}) p_t + (1-\beta_{fk_{ft}})(1-p_t)}. \tag{5}$$

Note that these updating rules do not depend on human capital h_t, conditional on job assignment. The reason is that the unobserved ability of a worker affects the probability that high or low output occurs, whereas human capital affects only the level of realized output and in the same way for both types of worker. Mechanically, $y_{fHk}(h_t)$ and $y_{fLk}(h_t)$ do not depend on α or β, but $\bar{y}_f(\alpha_{fk}, h_t, k)$ and $\bar{y}_f(\beta_{fk}, h_t, k)$ obviously do. Thus, human capital provides no additional information about the ability of a worker beyond the information that the realization of high or low output already conveys. Nonetheless, if high-ability workers produce high output more often than do low-ability workers in job k at firm f, that is, if $\alpha_{fk} > \beta_{fk}$, then a given amount of human capital, say, h_t, is allowed to differentially affect the expected output of the two types, as shown in the expression for $\bar{y}_f(\alpha_{fk}, h_t, k)$ in (3).

A key aspect of the model is that jobs can differ both in their expected output, as apparent from (2), and in their informativeness, since the probability of high output for a worker of high or low ability varies with the assigned job. In particular, in any period the job that yields (conditional on p_t and h_t) the highest current expected output may not be the most informative. Hence, in deciding on job assignment, firms not only take into account standard strategic considerations, as they would in any dynamic game of Bertrand wage competition, but they also take into account nontrivial dynamic learning considerations.

This interplay between output and information implies that firms trade off current expected

[5] In estimation I maintain that productivity shocks, as type I extreme value distributed, are unbounded. However, this specification is intended as an approximation to situations in which low realizations of productivity are possible.

output against greater information (and, possibly, human capital) when choosing a job for a worker. Thus, in this precise sense, firms face an *experimentation* problem when making their wage and job offers. By the same logic, workers also face an experimentation problem when choosing which wage and job offer to accept.

1.1.2 Production and Market Structure: Interpretation

Here I provide one interpretation of my specification of production in firms and market competition. Consider a set of primitive technologies, which represent different possible organizational structures of production within a firm. I allow for the possibility that two firms are *replicas* of each other, in that they have the same technology sets $\{y_{fHk}(h_t), y_{fLk}(h_t), \alpha_{fk}, \beta_{fk}\}_{k=1}^{K}$ as well as the same realizations of the productivity shocks $\boldsymbol{\varepsilon}_{ft}$.

I interpret different distributions of firms across production technologies as corresponding to different market structures, ranging from perfectly competitive (when each technology is adopted by many firms) to monopolistic (when each technology is adopted by one firm only). I think of imperfect competition for workers in the labor market as arising from a fixed cost of adopting any of the F possible technologies (or organizational structures). For instance, if technology f has a fixed cost $\kappa_f > 0$ of adoption, then the market can support the existence of, at most, one firm with this technology. The reason is that if two firms adopted the same technology, then competition between them would drive their profits to zero, so they would not be able to cover their fixed costs. In what follows, I will characterize equilibrium for any given distribution of firms across production technologies with this fixed cost setup providing one interpretation. I will then empirically investigate the degree of market power of the firm in my data.

1.1.3 Exogenous Separations

The model, as laid out so far, naturally generates endogenous separations, since firms with technologies better suited to a worker, given the prior p_t about the worker's ability, can successfully attract the worker. (In the Supplementary Appendix, I present an example that shows that, even without productivity shocks and human capital accumulation, mobility across jobs and firms occurs naturally in equilibrium.) In addition to endogenous separations, I allow for the possibility that a worker exogenously separates from a firm; that is, the worker may leave a firm for reasons unrelated to ability or performance. I assume that exogenous separations amount to a worker's permanent exit from the labor market considered.

Specifically, I assume that at the end of each period, a firm and a worker separate exogenously (an event denoted by $\zeta_{fkt} = 1$) with probability η_{fkt}, which depends on the firm f employing the worker, the job k the worker is assigned to, and the worker's labor market experience t. Such a shock captures instances in which the position the worker is assigned to closes, for instance, due to adverse demand conditions, or, alternatively, some preference shock induces the worker to leave the labor market under consideration. These shocks also reflect the random arrival of more attractive opportunities outside of the labor market I consider.

The assumption that exogenous separation shocks imply a worker's permanent exit from the market is natural given that, in my data, no worker who leaves the firm I study ever returns. Under this assumption, in the component game between all firms and a given worker, when an exogenous separation occurs, the game ends, and all firms and the worker obtain their reservation payoffs from then on. In practice, separation shocks imply that continuation values at t are discounted at rate $\delta(1-\eta_{fkt})$ if the worker is employed by firm f in job k in t.[6]

I restrict the separation rates η_{fkt} to depend on experience in the labor market only through a function of h_t, so from now on I will maintain that $\eta_{fkt} = \eta_{fk}(h_t)$. Hence, conditional on the firm f, the job k, and human capital h_t, these shocks do not depend on a worker's ability or performance. (In the empirical analysis, based on model diagnostic and fit, I restrict $\eta_{fkt}(h_t)$ to depend on just tenure at the currently employing firm.) From now on, I interpret $\eta_{fk}(h_t)$ as a further attribute of job technologies.

1.2 Timing

The timing of events in a period is as follows. First, at the beginning of any period t, each firm f draws a vector $\boldsymbol{\varepsilon}_{ft}$ of productivity shocks at each job. Then, all firms simultaneously submit their offers to workers. Each offer consists of a proposed wage, w_{ft}, and job assignment, k_{ft}, for a worker in the period. Next, each worker decides which offer to accept, the promised wage is paid, output at the accepted job at the employing firm is realized, and beliefs are updated based on observed output. Finally, exogenous separation shocks are realized.

I focus on the component game between all firms and one worker. In this game, the events in any period t are given by

$$(\boldsymbol{\varepsilon}_t, \mathbf{w}_t, \mathbf{k}_t, \mathbf{d}_t, y_{f_t k_{f_t t}}, \zeta_{f_t k_{f_t t}}) = \left(\{\boldsymbol{\varepsilon}_{ft}\}_f, \{(w_{ft}, k_{ft})\}_f, \{d_{ft}\}_f, y_{f_t k_{f_t t}}, \zeta_{f_t k_{f_t t}}\right),$$

where $\boldsymbol{\varepsilon}_t = \{\boldsymbol{\varepsilon}_{ft}\}_f$ denotes the vector of productivity shocks at each job of each firm; $(\mathbf{w}_t, \mathbf{k}_t) = \{w_{ft}, k_{ft}\}_f$, the vector of offers of each firm; $\mathbf{d}_t = \{d_{ft}\}_f$, the vector of the worker's decisions to accept ($d_{ft} = 1$) or reject ($d_{ft} = 0$) each firm's offer; $y_{f_t k_{f_t t}}$, the ability-dependent component of output realized in job $k_{f_t t}$ of the firm f_t employing the worker in the period; and $\zeta_{f_t k_{f_t t}}$, an indicator for whether a separation shock is realized at the end of the period in job $k_{f_t t}$ of firm f_t ($\zeta_{f_t k_{f_t t}} = 1$) or not ($\zeta_{f_t k_{f_t t}} = 0$).

[6] Of course, despite this exogenous exit rate, as long as new workers enter the market each period, there will always be workers in the market.

1.3 Equilibrium

Recall that I focus on the component game between all firms and a worker. I restrict attention to Markov perfect equilibria of this game in which the state variable at the beginning of any period t consists of the *informational human capital* of the worker, p_t, which is the common belief that the worker's ability is high, and the *technological human capital* of the worker, h_t, which summarizes the set of skills acquired by the worker with experience. Let $s_t = (p_t, h_t)$ denote the beginning-of-period state.

I now define the firms' and the worker's strategies in a Markov perfect equilibrium. I will refer to firm f's *offer* in a period as consisting of the proposed wage and job assignment, (w_{ft}, k_{ft}). The state that firms face at the time they make their offers is s_t, together with the vector of productivity shocks realized at each firm, $\boldsymbol{\varepsilon}_t$. The state that the worker faces when choosing which offer to accept consists of s_t, the productivity shocks realized at all firms, $\boldsymbol{\varepsilon}_t$, and the current vector of offers of all firms, $(\mathbf{w}_t, \mathbf{k}_t) = \big((w_{1t}, k_{1t}), \ldots, (w_{Ft}, k_{Ft})\big)$.

A *Markov perfect equilibrium* (MPE) consists of offer strategies $w_{ft} = w_f(s_t, \boldsymbol{\varepsilon}_t)$ and $k_{ft} = k_f(s_t, \boldsymbol{\varepsilon}_t)$ for firm $f \in \{1, \ldots, F\}$, an acceptance strategy for the worker, given by

$$\mathbf{d}(s_t, \boldsymbol{\varepsilon}_t, \mathbf{w}_t, \mathbf{k}_t) = (d_1(s_t, \boldsymbol{\varepsilon}_t, \mathbf{w}_t, \mathbf{k}_t), \ldots, d_F(s_t, \boldsymbol{\varepsilon}_t, \mathbf{w}_t, \mathbf{k}_t)) = (d_{1t}, \ldots, d_{Ft})$$

with $d_{ft} \in \{0, 1\}$, and updating rules $P_{fHk}(p_t)$ and $P_{fLk}(p_t)$ for all f and $k \in K^f$ such that the following holds:

(i) *Optimality by workers.* Given the strategies of all firms, the worker's strategy satisfies the Bellman equation

$$V^w(s_t, \boldsymbol{\varepsilon}_t, \mathbf{w}_t, \mathbf{k}_t) = \max_{\mathbf{d}} \Big\{ \sum\nolimits_f d_f \Big[(1-\delta) w_{ft} + \delta(1 - \eta_{fk_{ft}t}) \cdot \int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}, \mathbf{w}_{t+1}, \mathbf{k}_{t+1} | s_t, k_{ft}) dF(\boldsymbol{\varepsilon}_{t+1})] \Big\}, \tag{6}$$

where $\mathbf{d} = (d_1, \ldots, d_F)$ denotes the collection of the worker's possible acceptance ($d_f = 1$) or rejection ($d_f = 0$) decision with respect to each firm f's offer, $w_{ft} = w_f(s_t, \boldsymbol{\varepsilon}_t)$ and $k_{ft} = k_f(s_t, \boldsymbol{\varepsilon}_t)$, and

$$EV^w(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}, \mathbf{w}_{t+1}, \mathbf{k}_{t+1} | s_t, k_{ft}) = r_{k_{ft}}(p_t) V^w\left(P_{fHk}(p_t), h_{t+1}, \boldsymbol{\varepsilon}_{t+1}, \mathbf{w}_{t+1}, \mathbf{k}_{t+1}\right) + [1 - r_{k_{ft}}(p_t)] V^w\left(P_{fLk}(p_t), h_{t+1}, \boldsymbol{\varepsilon}_{t+1}, \mathbf{w}_{t+1}, \mathbf{k}_{t+1}\right)$$

denotes the worker's continuation value given the current prior p_t, the human capital h_t, and all firms' strategies. Here, $r_{k_{ft}}(p_t) = \alpha_{fk_{ft}} p_t + \beta_{fk_{ft}}(1 - p_t)$ denotes the probability of high output when the worker is assigned to job k_{ft} at firm f in period t, given the prior p_t.

(ii) *Optimality by firms.* Given the strategies of the other firms $g \neq f$ and the worker, the strategy of any firm f solves the Bellman equation

$$\Pi^f(s_t, \boldsymbol{\varepsilon}_t) = \max_{w,k}\Bigg(d_{ft}\Bigg\{(1-\delta)[y_f(s_t,k)+\varepsilon_{fkt}-w]+\delta(1-\eta_{fkt})\int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k)dF(\boldsymbol{\varepsilon}_{t+1})\Bigg\}$$

$$+\sum\nolimits_{g\neq f} d_{gt}\left\{\delta(1-\eta_{gk_{gt}t})\int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{gt})dF(\boldsymbol{\varepsilon}_{t+1})\right\}\Bigg), \qquad (7)$$

where $d_{ft} = d_f(s_t, \boldsymbol{\varepsilon}_t, \mathbf{w}_t, \mathbf{k}_t)$ is the acceptance decision of the worker with respect to firm f's offer, $k_{gt} = k_g(s_t, \boldsymbol{\varepsilon}_t)$ is part of the strategy of firm g,

$$E\Pi^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k) = r_k(p_t)\Pi^f\left(P_{fHk}(p_t), h_{t+1}, \boldsymbol{\varepsilon}_{t+1}\right) + [1-r_k(p_t)]\Pi^f\left(P_{fLk}(p_t), h_{t+1}, \boldsymbol{\varepsilon}_{t+1}\right),$$

and $E\Pi^f(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k_{gt})$ is defined analogously.

(iii) *Bayesian updating.* If p_t is the prior belief that the worker is of high ability at the beginning of period t and the worker's job assignment is k_{ft}, $f \in \{1, \ldots, F\}$, then if the output realized at the end of the period is high, the updated prior p_{t+1} at the beginning of period $t+1$ is given by (4). If, instead, the output realized at the end of the period is low, then the updated prior p_{t+1} at the beginning of period $t+1$ is given by (5).

In the following I restrict attention to *cautious Markov perfect equilibria* (MPEs), a subset of MPEs that result by imposing the restrictions that a firm not employing the worker in a period be indifferent between employing and not employing the worker and that such firm choose the job that maximizes the value of the match between that firm and the worker, which is the same way the winning firm chooses its offered job. (As Pastorino (2011) discusses, this notion is analogous to trembling-hand perfection and refines the notion of Bergemann and Välimäki (1996). See also DeVaro and Waldman (2012).)

This refinement implies that equilibrium is unique, since in a cautious MPE paid wages are uniquely determined. Moreover, in a cautious MPE the implied sharing rule of the surplus value generated by the employing firm and the worker depends on only the characteristics of the production sets of the employing firm and of the firm offering the worker the second-highest (expected present discounted) value of wages. Hence, this refinement allows for a flexible yet parsimonious way for surplus sharing to vary across workers and over time for the same worker, depending on the path of realized information and acquired human capital of a worker as summarized by s_t. I next characterize cautious MPEs and refer to them simply as *equilibria.*

2 Equilibrium Characterization

Here, I characterize the features of equilibrium that I will exploit in estimation: a reformulation of the problem defining the job choice of a given firm, the set of states at which that firm employs the worker, and the wage that firm pays when it employs. For notational convenience and in light of the empirical analysis, I focus on characterizing equilibrium from the point of view of just one firm, which I refer to as *firm A*.

Note that one approach to the estimation of the model is to directly recover its primitives based on the Bellman equation for each firm in (7). Ignoring constraints of data availability, however, proceeding this way is cumbersome because the solution to any given firm's profit maximization problem is a function of the strategies of all other firms as well as the strategies of the worker, each of which satisfies its own optimality conditions.

The approach I follow, motivated by data availability, exploits the restrictions on behavior that equilibrium implies to simplify the characterization of the outcomes of interest. Specifically, I rely on the equilibrium conditions that characterize the wage and job assignment offers of each firm to derive a joint surplus maximization problem, the *match surplus* problem, which combines the optimality conditions of a firm and the worker. The collection of the match surplus problem, evaluated at equilibrium, of each firm and the worker implies a direct mapping between primitive parameters and firm-level allocations that, by construction, subsumes all interdependence between the actions of the firms and the worker. I then use standard Bertrand logic to show that at any state only the second-best competitor (that is, the firm offering the second-highest expected present discounted value of wages) is relevant for any firm. I rely on this Bertrand logic, together with worker optimality, to derive a simple and intuitive expression for equilibrium wages. Finally, I impose a specializing assumption on the structure of the labor market to ensure that firms other than firm A behave competitively, even though firm A may not. This assumption allows me to simplify the match surplus problem of firm A so as to make it a convenient basis for estimation in light of my data.

The end result is that the equilibrium retention and job assignment problem of firm A reduces to one akin to a standard dynamic discrete choice problem, and equilibrium wages are given by a simple nonlinear equation in beliefs.

2.1 Key Propositions

Here, I set out my two key propositions. To this purpose, I define $V^A(\cdot) = \Pi^A(\cdot) + V^w(\cdot)$ to be the sum of the (expected present discounted) value of profits of firm A and the (expected present discounted) value of wages of the worker and refer to it as the *match surplus* value between firm A and the worker. Proposition 1 shows that this match surplus value solves a

Bellman equation that determines both the job offer of firm A and the set of states at which firm A employs the worker. In my empirical analysis, this match surplus value function constitutes one of the two fundamental estimating equations.

For convenience in stating the proposition, let $f = f(s_t, \boldsymbol{\varepsilon}_t)$ denote the firm, among the set of firms excluding firm A, whose job offer at t leads to the highest value of wages, given that the worker sequentially picks the best offer at each state. Given this definition of firm f, when firm A employs the worker, firm f is the second-best firm for the worker at $(s_t, \boldsymbol{\varepsilon}_t)$, and when firm A does not employ the worker, firm f is the first-best firm for the worker, that is, the employing firm. This notation makes it clear that firm f's identity changes with the state $(s_t, \boldsymbol{\varepsilon}_t)$ in that $f = f(s_t, \boldsymbol{\varepsilon}_t)$. Finally, let $k_{ft} = k_f(s_t, \boldsymbol{\varepsilon}_t)$ denote the job offered by firm $f = f(s_t, \boldsymbol{\varepsilon}_t)$ at state $(s_t, \boldsymbol{\varepsilon}_t)$.

Proposition 1. *In a cautious MPE, the match surplus value* $V^A(s_t, \boldsymbol{\varepsilon}_t)$ *equals*

$$\max\left\{\max_{k\in K^A}\left\{(1-\delta)[y_A(s_t,k)+\varepsilon_{Akt}]+\delta(1-\eta_{Akt})\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k)dF(\boldsymbol{\varepsilon}_{t+1})\right\},\right.$$

$$\left.(1-\delta)w_f(s_t,\boldsymbol{\varepsilon}_t)+\delta(1-\eta_{fk_{ft}t})\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{ft})dF(\boldsymbol{\varepsilon}_{t+1})\right\}, \tag{8}$$

where $f = f(s_t, \boldsymbol{\varepsilon}_t)$ *and* $k_{ft} = k_f(s_t, \boldsymbol{\varepsilon}_t)$*. Firm A employs the worker if, and only if,*

$$V^A(s_t,\boldsymbol{\varepsilon}_t)=\max_{k\in K^A}\left\{(1-\delta)[y_A(s_t,k)+\varepsilon_{Akt}]+\delta(1-\eta_{Akt})\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k)dF(\boldsymbol{\varepsilon}_{t+1})\right\}.$$

The proposition has three implications. First, regardless of whether firm A employs the worker at a particular equilibrium state or not, the sum of the values of firm A and the worker solves the same match surplus maximization problem. Second, the solution to this problem determines the set of states at which firm A employs the worker. Third, in equilibrium the job choice of firm A maximizes the sum of the values of firm A and the worker. Importantly, whereas this characterization of the job choice of firm A holds true in any MPE at states at which firm A employs the worker, it also applies to states at which firm A does not employ the worker by the cautious equilibrium restriction.

The next proposition derives the other fundamental equation that I use in estimation, namely, the expression for the wage (offered and) paid by firm A when it employs the worker.

Proposition 2. *In a cautious MPE, the worker's wage when employed by firm A is*

$$w_A(s_t,\boldsymbol{\varepsilon}_t)=y_f(s_t,k_{ft})+\varepsilon_{fk_{ft}t}+\frac{\delta}{1-\delta}\Psi_{ft}, \tag{9}$$

where the learning and human capital premium Ψ_{ft} *is given by*

$$\Psi_{ft} = \int_{\boldsymbol{\varepsilon}_{t+1}} \Big[(1-\eta_{fk_{ft}t})EV^f(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k_{ft}) - (1-\eta_{Ak_{At}t})EV^f(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k_{At})\Big]\, dF(\boldsymbol{\varepsilon}_{t+1}), \tag{10}$$

$f = f(s_t, \boldsymbol{\varepsilon}_t)$, $k_{ft} = k_f(s_t, \boldsymbol{\varepsilon}_t)$, *and* $k_{At} = k_A(s_t, \boldsymbol{\varepsilon}_t)$.

The expression for wages in (9) admits a natural interpretation. To help interpret it, first recall a static deterministic Bertrand game of price competition in which firms have asymmetric costs. In the standard equilibrium, the firm with the lowest cost sells the good and sets its price equal to the cost of the second-best firm, the firm with the second-lowest cost. Next, consider an analogous static Bertrand game in which firms have different production functions and compete in wages to attract workers. Similarly, in equilibrium the firm with the highest output employs the worker and sets its wage equal to the output of the second-best firm, the firm with the second-highest output.

In a special case of my model, an analogous result holds. Specifically, if three conditions are satisfied—if firm A and firm f have the same jobs, human capital evolves with tenure in the same way at the two firms, and productivity and separation shocks are perfectly correlated across the two firms—then the premium Ψ_{ft} is zero and the worker's wage reduces to

$$w_A(s_t, \boldsymbol{\varepsilon}_t) = y_f(s_t, k_{ft}) + \varepsilon_{fk_{ft}t}, \tag{11}$$

which is the exact analog of the static Bertrand pricing rule (in the presence of productivity shocks). According to (11), the winning firm pays a wage equal to the sum of the expected output of the worker at the second-best firm and the productivity shock in the job offered by that firm.

In general, the expression for paid wages is the sum of the current expected output of the worker at the second-best firm as in (11) and the premium Ψ_{ft}. This premium compensates the worker in t for the three dimensions along which jobs at the best and second-best firms differ and that affect the present value of wages from period $t+1$ on. These three dimensions reflect the fact that jobs at firm A and firm f may provide different amounts of information about the worker's ability, offer different prospects for human capital acquisition, and imply different separation shocks.

For example, if firm f's offered job is more informative than firm A's job, then if the worker chooses to work at firm f rather than at firm A in period t, the greater information acquired will affect the present value of the worker's wages from $t+1$ on. For another example, suppose that firm f has a technology that allows the worker to accumulate more human capital that is transferable across firms and, thus, raises the worker's present value of future wages. In both cases, when employed by firm A, the worker must be compensated for the forgone value of information or human capital. A similar logic applies to separation shocks.

2.2 Sketch of Proofs

Here, I sketch the proofs of these two propositions and leave the details to the Supplementary Appendix. I begin by discussing the immediate implications of worker and firm optimality from the definition of equilibrium, and then I use these implications to derive the match surplus maximization problem of firm A and the worker, and the expression for paid wages. To simplify notation, I denote the job offer of any firm g different from A as $k_{gt} = k_g(s_t, \boldsymbol{\varepsilon}_t)$.

Consider first the worker's problem. Optimality by the worker in (6) immediately implies that the worker chooses the firm that offers the combination of current wage and job leading to the highest present value of wages. Thus, if the worker chooses firm A, it must be that

$$(1-\delta)w_A(s_t, \boldsymbol{\varepsilon}_t) + \delta(1-\eta_{Ak_{At}t}) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k_{At})dF(\boldsymbol{\varepsilon}_{t+1})$$

$$\geq (1-\delta)w_g(s_t, \boldsymbol{\varepsilon}_t) + \delta(1-\eta_{gk_{gt}t}) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k_{gt})dF(\boldsymbol{\varepsilon}_{t+1}) \tag{12}$$

for any other firm g. Note that (12) holds with equality when the other firm is the second-best firm f. In (12) I have suppressed the dependence on firms' future offers in $EV^w(\cdot)$. To be precise, in equilibrium, wage and job offers by all firms, $\mathbf{w}_t(s_t, \boldsymbol{\varepsilon}_t)$ and $\mathbf{k}_t(s_t, \boldsymbol{\varepsilon}_t)$, depend on $(s_t, \boldsymbol{\varepsilon}_t)$, so the value function of the worker is $V^w(s_t, \boldsymbol{\varepsilon}_t, \mathbf{w}_t(s_t, \boldsymbol{\varepsilon}_t), \mathbf{k}_t(s_t, \boldsymbol{\varepsilon}_t))$, which can be expressed more compactly, by a slight abuse of notation, as $V^w(s_t, \boldsymbol{\varepsilon}_t)$. Thus, in equilibrium, the match surplus value $V^f(\cdot) = \Pi^f(\cdot) + V^w(\cdot)$ also depends on only $(s_t, \boldsymbol{\varepsilon}_t)$.

Consider now the employing firm's problem. Firm A realizes that the worker will accept its wage and job offer only if that offer (w_{At}, k_{At}) is at least as attractive as any other firm's. Thus, when firm A employs, it maximizes the value of its profits subject to the constraint (12). That is, firm A solves this problem:

$$\max_{w,k} \left\{ (1-\delta)[y_A(s_t, k) + \varepsilon_{Akt} - w] + \delta(1-\eta_{Akt}) \int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k)dF(\boldsymbol{\varepsilon}_{t+1}) \right\} \tag{13}$$

subject to (12).

Clearly, firm A chooses its offer so that (12) holds as an equality when evaluated at the offer of the second-best firm; otherwise firm A could increase its profits by modifying its offer. This logic follows the standard argument for a static Bertrand pricing game when firms have different costs. Since the constraint (12) holds as equality at $f = f(s_t, \boldsymbol{\varepsilon}_t)$, I can use (12) to substitute out the wage offer of firm A from (13). Dropping irrelevant constants independent of firm A's choices, I obtain that firm A's job offer when it employs, $k^A(s_t, \boldsymbol{\varepsilon}_t)$, equals

$$\arg\max_{k \in K^A} \left\{ (1-\delta)[y_A(s_t,k) + \varepsilon_{Akt}] + \delta(1-\eta_{Akt}) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k)dF(\boldsymbol{\varepsilon}_{t+1}) \right\}, \quad (14)$$

where, by definition, $V^A(s_t, \boldsymbol{\varepsilon}_t) = \Pi^A(s_t, \boldsymbol{\varepsilon}_t) + V^w(s_t, \boldsymbol{\varepsilon}_t)$. The intuition here is that firm A realizes that by offering a job that increases the worker's continuation value, it can cut back on its current wage offer one-for-one. Since this wage enters the same way in the constraint and in the objective function, the firm has an incentive to choose a job in order to maximize the sum of its profits and the worker's value.[7]

Next, when firm A employs the worker, it must earn a higher present value of profits than if it does not employ and, instead, the worker works at firm f. If the worker was employed by firm f, firm A would earn zero profits in the current period, but it would update its beliefs about the worker's ability according to the output realized at job k_{ft} of firm f. Thus, at states at which firm A employs the worker, it must be that

$$\Pi^A(s_t, \boldsymbol{\varepsilon}_t) = (1-\delta)[y_A(s_t, k_{At}) + \varepsilon_{Ak_{At}t} - w_{At}] + \delta(1-\eta_{Ak_{At}t})$$

$$\cdot \int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k_{At})dF(\boldsymbol{\varepsilon}_{t+1}) \geq \delta(1-\eta_{fk_{ft}t}) \int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k_{ft})dF(\boldsymbol{\varepsilon}_{t+1}), \quad (15)$$

where $k_{At} = k_A(s_t, \boldsymbol{\varepsilon}_t)$, $w_{At} = w_A(s_t, \boldsymbol{\varepsilon}_t)$, and $k_{ft} = k_f(s_t, \boldsymbol{\varepsilon}_t)$. Likewise, at equilibrium states at which firm A does not employ the worker, firm A must weakly prefer to lose the worker to firm f. So

$$\Pi^A(s_t, \boldsymbol{\varepsilon}_t) = \delta(1-\eta_{fk_{ft}t}) \int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k_{ft})dF(\boldsymbol{\varepsilon}_{t+1}) \geq$$

$$(1-\delta)[y_A(s_t, k_{At}) + \varepsilon_{Ak_{At}t} - w_{At}] + \delta(1-\eta_{Ak_{At}t}) \int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k_{At})dF(\boldsymbol{\varepsilon}_{t+1}) \quad (16)$$

when firm $f = f(s_t, \boldsymbol{\varepsilon}_t)$ employs the worker at job $k_{ft} = k_f(s_t, \boldsymbol{\varepsilon}_t)$ and firm A makes a losing offer of $w_{At} = w_A(s_t, \boldsymbol{\varepsilon}_t)$ and $k_{At} = k_A(s_t, \boldsymbol{\varepsilon}_t)$.

Notice that condition (16) does not uniquely pin down a losing firm's wages in an MPE. Recall that the winning firm's wages are pinned down by the constraint (12) holding as an equality against the second-best firm. So when (16) holds as an inequality for the second-best firm, the second-best firm's offer is not pinned down and, in turn, the winning firm's wages are not uniquely determined either. The cautious restriction resolves this indeterminacy. Specifically, from the requirement that a losing firm be indifferent between employing and not

[7] Briefly, the constraint holds with equality in a problem of this form: choose x and w to solve $\max_{w,x}[f(x) - w]$ subject to $w + g(x) \geq M$. Substituting out for w reduces the problem to this form: choose x to solve $\max_x[f(x) + g(x) - M]$ where M is an irrelevant constant that can be dropped.

employing the worker—so that (16) holds as an equality—the wage of a losing firm is uniquely pinned down and, correspondingly, the wage of the employing firm is also uniquely determined from (12).

Now conditions (12)–(16) can be manipulated to derive (8), (9), and (10). Briefly, (8) follows by adding the worker's value to firm A's value both when firm A employs the worker and when it does not. The expressions for wages, (9) and (10), follow by imposing that (16) holds as an equality, where now firm f is the losing firm, solving for the losing firm's wage, and substituting the resulting expression into (12) holding as an equality against firm f to obtain the wage of firm A, the winning firm. See the Supplementary Appendix for further details.

2.3 Specialization: Perfect Competition Among Competitors

In estimation, again, I use data for one particular U.S. firm, which I refer to as *my firm*. In the following, I maintain a simple assumption on the competitiveness of other firms in the same labor market in which the firm in my data hires its managers. Namely, since I have no direct information about other firms, I assume that these competitors of my firm set wages in a perfectly competitive way. This assumption allows me to focus on the strategic considerations of my firm while keeping the theoretical analysis and estimation tractable.[8]

I now discuss in some detail the assumptions on technologies that lead other firms to set their wages as under perfect competition. I assume that for each technology that is adopted by some firm f other than my firm, firm A, at least one other firm adopts the same technology. Under this assumption, the competition between these replica firms allows the worker to extract the full surplus from a match with any such firm, leaving all of them with zero profits. Thus, if some firm adopts a technology set K^f, with productivity shocks $(\varepsilon_{f1t}, \ldots, \varepsilon_{fK^ft})$ and separation shocks $(\eta_{f1t}, \ldots, \eta_{fK^ft})$ at time t, then so does at least one other firm. (One interpretation of the fact that shocks are perfectly correlated across replica firms is that they are an attribute of the job technology.) Hence, whenever one of these firms employs the worker, the second-best firm will be a replica of it. These two firms will make the same job and wage offer, with jobs characterized by the same information content and skill acquisition possibilities. Hence, the compensating premium Ψ_{ft} in (10) between these two firms with technology set K^f is zero. As a result, when one of these firms f with technology set K^f employs the worker, it will offer and pay the wage

$$w_f(s_t, \boldsymbol{\varepsilon}_t) = y_f(s_t, k_{ft}) + \varepsilon_{fk_{ft}t}. \tag{17}$$

When, instead, a firm with technology f does not employ the worker, its (job and) wage offer must make it indifferent between employing and not employing the worker by the cautious equilibrium restriction. Since its profits are zero, by the cautious restriction this firm's wages

[8] I thank one of the referees for suggesting this formulation.

are also given by (17). To see this result, consider the expression for the wage of a losing firm derived from making (16) hold as an equality and replace A with f as the losing firm. Then (17) follows from the fact that firm f's continuation profits must be zero.

I have already shown that firm A solves (8). Combining (8) with the result in (17), the match surplus value function, $V^A(s_t, \boldsymbol{\varepsilon}_t)$, in (8) specializes to

$$\max\left\{\max_{k\in K^A}\left\{(1-\delta)[y_A(s_t,k)+\varepsilon_{Akt}]+\delta(1-\eta_{Akt})\int_{\boldsymbol{\varepsilon}_{t+1}}EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k)dF(\boldsymbol{\varepsilon}_{t+1})\right\},\right.$$

$$\left.(1-\delta)[y_f(s_t,k_{ft})+\varepsilon_{fk_{ft}t}]+\delta(1-\eta_{fk_{ft}t})\int_{\boldsymbol{\varepsilon}_{t+1}}EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{ft})dF(\boldsymbol{\varepsilon}_{t+1})\right\}, \qquad (18)$$

where $k_{ft} = k_f(s_t, \boldsymbol{\varepsilon}_t)$. Observe that here the best-response job assignment problem of firm A evaluated at equilibrium is analogous to a standard single-agent dynamic discrete choice problem among $K^A + 1$ alternatives. The difference compared to the standard problem is that the alternative to one of the K^A jobs of firm A is determined by equilibrium. These observations lead to my third proposition.

Proposition 3. *Consider firm A. Under the assumption that, for any technology used by a competing firm, at least one other firm has access to the same technology, in a cautious MPE the match surplus value function between firm A and the worker simplifies to (18).*

Under this replica assumption, the model is equivalent to one in which there is only one firm for each technology set, other than the technology set of firm A, and each of these other firms prices competitively, in that each firm sets wages according to (17). Of course, it is also equivalent to one in which all of the outside firms are replaced by a single outside firm that has access to all the technologies of the outside firms and sets wages competitively.

Nonetheless, this setup allows the firm I study to have market power and obtain positive profits. Note for later that I can indirectly infer the extent to which my firm has market power by testing the restriction that my firm pays wages equal to the worker's expected output at the firm; that is, $w_A(s_t, \boldsymbol{\varepsilon}_t) = y_A(s_t, k_{At}) + \varepsilon_{Ak_{At}t}$.

3 Data

Now I describe the data I use, present some descriptive statistics from those data, and explain intuitively how my model can account for them. In the Appendix, I provide more details about the estimation sample. In the Supplementary Appendix, I describe the construction of the estimation sample and contrast it to a larger sample on which I reestimate the model, obtaining similar results to those presented here. See the Supplementary Appendix for details.

3.1 The Firm and the Job Hierarchy

The source of my data, collected and first analyzed by BGH, is the personnel records of all management employees of a medium-sized U.S. firm in a service industry between 1969 and 1988. As described by BGH, these records include information on every managerial employee in the firm as of December 31 of each year. Each record contains for each manager an employee's identification number, year of entry, age, education, job title and level, cost center code (the six-digit code of the organizational unit defined for measuring costs, revenues, or profits), salary, salary grade (available from 1979 to 1988), bonus, and job performance rating (from 1, highest, to 5, lowest). As for performance, the precise timing of ratings is unknown; according to BGH, ratings are probably year-end values, since bonuses each year are known to be awarded on February 1 of the following year. Similarly, whereas title and salary are year-end values, it is unclear precisely when changes in these values occurred in a year. To be consistent with the model, in the empirical analysis I assume that title and pay changes occur at the end of the year *after* individual performance is appraised and that these changes apply to the subsequent year. In the rest of the paper, I refer to salary or base pay simply as *wage*.

Over the 20-year sample period, the firm has been remarkably stable in the composition of titles and levels of the job hierarchy. Even as firm size has tripled, the fraction of managerial employees at each level has not significantly changed. After 1984 some new titles were created, but only two are of significant size, representing only 0.6 percent and 0.9 percent of employees. (See BGH's Table I.) Titles were not coded for some new hires in later years. Specifically, missing data are significant in 1987 and 1988, when approximately 10 percent of employees and half of new hires do not have title information.

The size of entry cohorts into managerial positions grew significantly during the sample period. The entry cohort had 230 individuals in 1970 and 1,175 by 1988. The exit rate from managerial ranks is substantial in each year. For instance, for the group of entrants at the firm between 1970 and 1979, 10.9 percent had left the firm after 1 year, 20.4 percent after 2 years, and 65.9 percent after 10 years. As noted by BGH, career patterns are similar across early and later entrants even as the average career length shortens.

The data include demographic characteristics for the firm's managers. The average age of managers is 39 years with a standard deviation of approximately 10 years, from a minimum of 20 to a maximum of 71. Their average number of years of education is 15, with a standard deviation of approximately 2 years, from a minimum of 12 (high school degree) to a maximum of 23 (Ph.D.). Both age and education display little variation across cohorts. The composition of entrants across job titles does not change markedly over the years either, though lower-level entry increases between 1976 and 1985.[9]

[9]Baker, Gibbs, and Holmström (1994b) report that the proportion of minorities and women at this firm increased steadily. My copy of the data does not include information on gender or race.

The firm's internal hierarchy of jobs constructed by BGH consists of two clearly distinct parts, Levels 1–4 (the lower ranks) and Levels 5–8 (the upper ranks), where Level 8 corresponds to the highest title of Chairman-CEO. (For an extensive discussion of the procedure employed to aggregate job titles into levels, see BGH.) Levels 1–4, each of approximately the same size, employ 97.6 percent of managers. Over the sample period, 16,981 employees are at Level 1, 17,725 at Level 2, 17,253 at Level 3, and 13,892 at Level 4. The corresponding figures at Levels 5-8 are 1,194, 373, 56, and 20. Note that level information is missing for 6,577 employees over the sample years.

In estimation I focus on 1,426 managers who entered managerial positions between 1970 and 1979 at Level 1 with at least 16 years of education at entry, who experienced no change in the recorded number of years of education during the first 10 years of tenure at the firm, and have no level information missing during their first 10 years at the firm. Due to the high separation rate in each year and at each tenure, I restrict attention to the first 8 years of observations on these managers. In addition, given the small number of individuals observed at Levels 3 and higher, I treat observations on such levels as observations on Level 3. Then Level 1, Level 2, and this composite Level 3 correspond to job 1, job 2, and job 3 in the model.

Restricting attention to entrants at only one level allows for a parsimonious parameterization of initial beliefs and human capital, which, for instance, do not need to be interacted with the entry level of a manager. Once the focus of the analysis is on entrants at only one level, the obvious candidate level is Level 1, because the vast majority of entrants (over 70 percent) enter at this level. To address potential concerns about sample selection, I nonetheless reestimated the model on a larger sample that includes entrants at Levels 2, 3, and 4 as well. It turns out that the estimates of the model's parameters based on this larger sample are remarkably similar to those based on the smaller sample I focus on here. See the Supplementary Appendix for details.

Finally, even though performance ratings range from 1 (highest) to 5 (lowest), in the data ratings of 3, 4, and 5 are only a small fraction of all ratings. Specifically, over the first eight years of tenure in my subsample, ratings of 1 and 2 account for 96.3 percent, 92.3 percent, 92.5 percent, 93.2 percent, 89.7 percent, 86.8 percent, 86.1 percent, and 83.2 percent of all rating observations. Thus, I combine ratings from 2 to 5 into a single measure to obtain a simple binary classification of high and low performance, as in the model. According to this reclassification, in my sample a rating of 1 corresponds to high performance and a rating of 0 to low performance.

3.2 Descriptive Statistics: Presentation and Interpretation

Here, I present a few descriptive statistics calculated from the estimation sample concerning the distribution of managers across levels by tenure, the level distribution of performance ratings by tenure, and the level distribution of wages by tenure. I summarize the key features of these statistics and argue, intuitively, how the model can account for them.

3.2.1 Level Assignments

For the estimation sample, Table 1 displays the percentage of managers at each level, as well as the percentage of managers who separate from the firm in each of the first seven years of tenure. (In estimation I only used partial information on observed outcomes in the eighth year of tenure. See the derivation of the likelihood function in the Supplementary Appendix for details.) Overall, the percentage of employees at Level 1 and Level 2 rapidly decreases with tenure at the firm, whereas the percentage of employees assigned to Level 2 and Level 3 first increases (respectively, until the third and the sixth year of tenure) and then decreases. In addition, the percentage of managers who separate from the firm is substantial at each tenure. By the seventh year, 57.6 percent of those hired at Level 1 have left the firm.

For ease of interpretation, I convert the level distribution of managers in Table 1 into level-specific hazard rates of separation, retention at the same level, and promotion to the next level by tenure, as shown in Table 2. For example, from tenure 1 to 2, 14.5 percent of managers separate, 45.6 percent are retained at Level 1, and 39.9 percent are promoted to Level 2. Note that in the sample no manager is ever demoted, and all promotions are by just one level.

Table 2 shows that at each level, the hazard rates of separation are approximately constant in tenure. The hazard rates of promotion both from Level 1 to Level 2 and from Level 2 to Level 3 increase initially, then essentially decrease. For promotion from Level 1 to Level 2, the increase occurs between the first and second year of tenure, and it is sizeable (from 39.9 percent to 48.6 percent). Then, the hazard rate essentially declines. For promotions from Level 2 to Level 3, this increase occurs between the second and third year of tenure, and it is also sizeable (from 21.8 percent to 37.4 percent). Then, the hazard rate declines.

Now consider how my model can explain this nonmonotone tenure profile of the hazard rate of promotion. Without heterogeneous initial priors or technological human capital acquisition or productivity and separation shocks, a *pure learning* version of my model can naturally explain why the hazard rate of promotion decreases with tenure. Simply put, the best managers will tend to experience more successes, hence their priors will increase, and they will be promoted if higher ability is more valuable at higher levels. As more managers are promoted, the average ability of the pool of unpromoted managers worsens over time, everything else equal, implying a decreasing probability of promotion.

The presence of heterogeneous initial priors can produce the initial increase in the rate of promotion. To see how, consider the following example. If a modest percentage of managers have relatively high priors at a level, then a single success will lead, on average, to their promotion to the next level. This group of managers—call it group 2—will account for the initial hazard rate of promotion. But now suppose that another group of managers—call it group 1—both is larger in size and consists of managers who have smaller priors than those in group 2. Such managers may need, for instance, two successes to reach the cutoff level for promotion. In the second year of tenure, the mass of these group 1 managers with two consecutive successes can easily be large enough that the hazard rate of promotion increases. After these two initial periods, the selection effect discussed can then lead to a decreasing hazard rate of promotion. (Since a promotion is typically associated with a wage increase, this mechanism also explains why managers who receive large wage increases early in their stay at one level of the job ladder are promoted quickly to the next.)

In the data real wage decreases are frequent, but demotions are never observed. The pure learning version of the model has a difficult time in generating both of these observations simultaneously. In contrast, once I augment that model to include technological human capital, it can generate both. (Real wage decreases occur naturally, since poor performance lowers the firm's assessment of a manager's ability and, by competition, the manager's wage; see the discussion below.) Real wage decreases are not accompanied by demotion because of technological human capital accumulation. In particular, if a manager acquires enough technological human capital, either task-specific or transferable to higher levels, then it is not in the firm's best interest to demote the manager even in the face of poor performance. Rather, retaining the manager at the same level is optimal. See a similar discussion in GW.

Finally, the relatively constant probability of separation can be due to an asymmetric signal structure across high- and low-ability managers, when, for instance, just one low performance rating causes the prior to decrease to almost zero, since the probability of a single failure is constant. (Omitting the firm subscript, this would be the case if $\alpha_k = 1$, $\beta_k \neq 1-\alpha_k$, and $\beta_k > 0$ at some job k of firm A. In this case the prior would converge to zero after a single failure at job k. See also the Supplementary Appendix.) More simply, the relatively constant probability of separation can be due to a relatively constant probability of exogenous separation.

3.2.2 Performance Ratings

The performance ratings of the managers at my firm are displayed in Table 3. The second, third, and fourth columns show the percentage of all managers at Levels 1, 2, and 3 who receive a high rating in tenure t. The next two columns display, in each tenure t, the percentage of high ratings among those managers who are assigned to Level 1 in t and are retained at Level

1 in $t + 1$ and the percentage of high ratings among those managers who are assigned to Level 1 in t and are promoted to Level 2 in $t + 1$. The last two columns display the corresponding percentages for managers who are assigned to Level 2 in t and are retained at Level 2 in $t + 1$ and those who are assigned to Level 2 in t and are promoted to Level 3 in $t + 1$.

The patterns of high ratings display four distinctive features. First, at any given level, the percentage of high ratings at the level decreases with tenure (apart from $t = 6$ at Level 1 and $t = 7$ at Level 2). Second, at any given tenure, the percentage of high ratings increases with the level. Third, the percentage of high ratings among promoted managers is higher than among unpromoted managers (apart from $t = 5$ at Level 1 and $t = 7$ at Level 2). Fourth, the percentage of high ratings among promoted managers is higher among those managers promoted early in their tenure at a level relative to those promoted later in their tenure at a level (apart from $t = 6$ at Level 1).

Now consider how my model can explain these patterns. Again, without heterogeneous initial priors or technological human capital acquisition or productivity and separation shocks, a pure learning version of my model can naturally explain the first three features. The core mechanism of the learning model is that successful performance, as measured by a high rating, increases the perceived ability of a manager by increasing the prior and, thus, leads to promotion (if high ability is more valuable at high-level jobs). This mechanism obviously generates the third feature, that promoted managers receive higher ratings. As a by-product, this mechanism generates the first and second features as well. If managers with higher perceived abilities are promoted, then the model naturally produces the second feature: managers who make it to higher levels have higher actual abilities and, hence, receive higher ratings. Finally, the model can also produce the first feature by a selection effect: as the better managers are promoted, the ability of the pool of unpromoted managers worsens with tenure.

The presence of heterogeneous initial priors can produce the fourth feature, that managers promoted early tend to receive higher ratings than those promoted later. Recall the earlier example with two groups of managers, where managers in group 2 have higher initial priors at entry, say, p_{21}, than those in group 1, say, p_{11}. (The first subscript denotes the manager group, the second subscript the year of tenure.) The key to generating this feature is that managers in group 1 are promoted *after* managers in group 2 and that the prior at promotion for group 1 is *smaller* than the prior at promotion for group 2.

To see how, consider the following example. Dropping the subscript denoting the firm for simplicity, note that the expected fraction of high ratings at Level k at the end of the first year of tenure equals $\alpha_k p_{i1} + \beta_k (1 - p_{i1})$ for group i. Suppose that group 2 managers need one success to be promoted and that group 1 managers need two successes. Suppose also, critically, that after one success, the posterior about group 1 managers, $P_{Hk}(p_{11})$, is smaller than the initial prior of group 2 managers, p_{21}. Then, the fraction of high ratings among those promoted at the end of the second year of tenure (group 1 managers with a success in the first year, who experience a second success the next year and are promoted) will be lower than the fraction of high ratings among those promoted at the end of the first year of tenure (group 2 managers, who succeed in the first year and are promoted), since $\alpha_k P_{Hk}(p_{11}) + \beta_k[1 - P_{Hk}(p_{11})] < \alpha_k p_{21} + \beta_k(1 - p_{21})$ with $\alpha_k > \beta_k$ and $P_{Hk}(p_{11}) < p_{21}$.

I have focused on a simple example with two groups of managers, but in the empirical analysis, I allow for more than two. Depending on the relative initial priors and sizes of these groups, the model can generate rich patterns for the timing of promotions and the distribution of ratings among promoted and unpromoted managers at different tenures. Conversely, the rich patterns observed in the data regarding these features of job mobility provide information about the number and initial priors of these unobserved groups of managers. (See the Supplementary Appendix for a detailed discussion of model identification.)

3.2.3 Wages

Tables 4 and 5 display the distribution of wages by level and tenure, and the statistics on the distribution of wage changes by tenure. In Table 4, the second, third, and fourth columns display the percentage of managers who earn between \$20–40 thousand (1988 U.S. dollars), \$40–60 thousand, and \$60–80 thousand at each level by tenure. The last two columns display the mean and standard deviation of wages by level and tenure.

Three features are apparent in Table 4. First, comparing mean wages at the three levels reveals that mean wages are increasing in the level. (Indeed, the mean wage, pooled across tenures, is \$39,382 at Level 1, \$43,433 at Level 2, and \$50,351 at Level 3.) Second, the standard deviation of wages at almost all tenures is higher at Levels 2 and 3 than at Level 1. (The standard deviation of wages, pooled across tenures, is \$6,924 at Level 1, \$7,377 at Level 2, and \$7,270 at Level 3.) Third, the mean wage at a given level decreases with tenure at Level 1, whereas it is approximately constant at the higher levels.

Two additional features are apparent in Table 5. The second, third, and fourth columns of this table display the percentage of negative wage changes, positive wage changes smaller than 15 percent, and positive wage changes between 15 and 30 percent by tenure, pooled across levels. The last column reports the yearly growth rate of mean wages by tenure. This table provides a fourth feature of wages in the data: at each tenure, over 20 percent of wage changes are negative. A fifth feature is that although mean wages increase with tenure, the yearly growth rate oscillates.

Consider how, for instance, the pure learning version of my model, without technological human capital and without productivity or separation shocks, can produce these patterns. The key to generating all of them are two assumptions: that a manager's ability and prior at my firm are correlated with the manager's ability, and hence output, and prior at other firms and that firms in the labor market compete for managers. These two assumptions imply that perceived ability at my firm is correlated with wages, so increases in perceived ability lead, on average, to increases in wages. Given these two assumptions, the pure learning version of my model naturally generates all five features.

First, wages increase with the level if, as discussed earlier, the perceived ability of a retained manager increases with the level. In this case, competition between my firm and other firms in the market leads to wage increases as perceived ability increases.

Second, as managers spend more time in the firm, their performance outcomes become more heterogeneous. For instance, 2^{t-1} histories of successes and failures are possible in the t-th period of tenure. So the variance of wages naturally increases over time and, if managers are assigned to higher levels over time as information about them accumulates, then the variance of wages also increases with the level.

Third, if better managers are promoted out of Level 1, as discussed earlier, then the prior of unpromoted managers falls; hence, their wages on average decrease. This selection effect is smaller at Level 2 and absent at Level 3, since at Level 3 there is no possibility of further promotion. This mechanism then provides an explanation for why the mean wage decreases with tenure primarily at Level 1.

Fourth, the learning model naturally generates wage decreases after failures, as discussed.

Finally, the model allows for two sources of growth in average wages with tenure. One source is that the model generates an increase in the assigned level for managers retained by the firm; that is, managers who are retained are progressively promoted from Level 1 to some combination of Levels 2 and 3 over time. Since average wages increase with the level, overall average wages increase with tenure. This increase can be generated either by learning, if the prior increases with the level, or by human capital acquired at the firm, which is transferable across levels and firms. The other possible source of wage growth with tenure is related. If only managers who are perceived to be sufficiently able and productive are retained, then the selection effect of retention implies that the average ability of retained managers and, hence, their average wages increase with tenure. The fact that the growth rate of wages oscillates over tenure may be due, once more, to a composition effect from the varying fraction of managers with different priors at different tenures and, thus, to the varying fraction of managers experiencing high performance and wage increases.

In light of this evidence that the model is broadly consistent with the main patterns of interest in the data, I now turn to evaluating more formally its ability to replicate them.

4 Empirical Specification

Here I describe the empirical specification of the model and the derivation of the main objects entering its likelihood function: the probabilities of job assignment and separation, the probabilities of performance ratings, and the probabilities of paid wages. Omitted details are collected in the Supplementary Appendix. Then I provide an overview of the identification of the model, relegating an extended and more formal discussion to the Supplementary Appendix. A detailed description of the model specification, as well as the numerical solution and estimation procedure I employ, are contained in the accompanying Appendix and Supplementary Appendix. Since in my data I observe managers continuously employed at one firm, the subscript t will now denote tenure at my firm.

4.1 Human Capital and Unobserved Heterogeneity

I first describe my specification of the process of technological human capital acquisition and of the other dimensions of unobserved heterogeneity among managers, besides ability.

Technological Human Capital. Formally, denote my firm as firm A and recall that $s_t = (p_t, h_t)$. In my data, managers are continuously employed at one firm only. Therefore, the amount of new skills that any manager can accumulate over the sample years depends, at most, on the manager's tenure at each of the firm's jobs. In practice, the human capital acquired by a manager at the firm turned out to depend on only the manager's tenure at my firm, $t-1$, and the previous period job assignment, k_{t-1}, conditional on the initial human capital, h_1. Given the interpretation of t as current tenure at my firm, I have correspondingly interpreted h_1 as human capital acquired by a manager before entry into my firm.

Empirically, h_1 is captured by the only observed characteristics of a manager at entry into the firm, namely, a manager's age, education, and year of entry. Hence, (1) reduces to $h_t = h(h_1, t-1, k_{t-1})$ for $t \geq 2$, where k_{t-1} denotes a manager's assignment in tenure $t-1$. (In the specification for paid wages, h_1 will also be indexed by n, to express its dependence on the vector $e_n = (age_n, edu_n, year_n)$ of observed characteristics of a manager at entry, a manager's age (age_n), education (edu_n), and year of entry ($year_n$).)

I flexibly parameterize high and low output at each job k as $y_{AHk}(h_t) = y_{AHk} + A_k(h_t) + B_k(h_t)$ and $y_{ALk}(h_t) = y_{ALk} + B_k(h_t)$. By simple algebra, from (3) it follows that the expected output of a manager known to be of high ability is

$$\bar{y}_A(\alpha_{Ak}, h_t, k) = \alpha_{Ak} y_{AHk} + (1-\alpha_{Ak}) y_{ALk} + \alpha_{Ak} A_k(h_t) + B_k(h_t), \tag{19}$$

whereas the expected output of a manager known to be of low ability is

$$\bar{y}_A(\beta_{Ak}, h_t, k) = \beta_{Ak} y_{AHk} + (1 - \beta_{Ak}) y_{ALk} + \beta_{Ak} A_k(h_t) + B_k(h_t). \tag{20}$$

Empirically it turned out that, conditional on the current job assignment k, $A_k(h_t)$ reduces to A_{kt}, that is, it depends on only current job assignment and tenure. Similarly, $B_k(h_t)$ depends on only current job assignment, tenure, and previous period assignment. Therefore, I model $B_k(h_t)$ as

$$B_k(h_t) = \sum\nolimits_{k'=1}^{3} b_{kk_{t-1}t} I(k_{t-1} = k'),$$

where $I(k_{t-1} = k')$ is an indicator function that equals one if, and only if, a manager's previous period assignment is k'. Note that a large positive value for $b_{kk_{t-1}t}$ when $k = k_{t-1}$ and a small value when $k \neq k_{t-1}$ imply a large degree of task-specificity of the skills acquired at level k_{t-1} in $t-1$. Under these specializations, (19) becomes

$$\bar{y}_A(\alpha_{Ak}, h_t, k) = \alpha_{Ak} y_{AHk} + (1 - \alpha_{Ak}) y_{ALk} + \alpha_{Ak} A_{kt} + \sum\nolimits_{k'=1}^{3} b_{kk_{t-1}t} I(k_{t-1} = k'). \tag{21}$$

A similar expression for $\bar{y}_A(\beta_{Ak}, h_t, k)$ can be obtained by replacing α_{Ak} in (21) with β_{Ak}. By (2), $y_A(p_t, h_t, k)$ can now be expressed as

$$y_A(p_t, t-1, k_{t-1}, k) = a_{kt} + \sum\nolimits_{k'=1}^{3} b_{kk_{t-1}t} I(k_{t-1} = k') + c_{kt} p_t, \tag{22}$$

where the parameters $a_{kt} = \beta_{Ak} y_{AHk} + (1 - \beta_{Ak}) y_{ALk} + \beta_{Ak} A_{kt}$ and $c_{kt} = (\alpha_{Ak} - \beta_{Ak})(y_{AHk} - y_{ALk} + A_{kt})$ are constants that depend on only a manager's current assignment and tenure at the firm. In a slight abuse of terminology, from now on I will refer to a_{kt}, $b_{kk_{t-1}t}$, and c_{kt} as *productivity and technological human capital parameters.*

Notice that in (22) the impact of h_1 on a manager's expected output at my firm is subsumed by the other state variables. The reason is that, in practice, observed job assignments do not display significant variation with a manager's observed characteristics at entry, conditional on the manager's other state variables. Nonetheless, I allow h_1 to affect a manager's productivity at other firms and thus, due to competition in the labor market, a manager's wage at my firm. Specifically, the part of h_1 that is transferable to my firm is captured by the initial prior p_1 (with $\Pr(p_1|p_0, k_0) = \Pr(p_1)$). The part that is transferable to other firms, instead, is captured by the observed characteristics of a manager at entry into the firm, that is, a manager's entering age, entering education, and year of entry, which affect paid wages.

Recall that here I focus on the sample of managers entering into the firm at Level 1. I also estimate a version of the model on a larger sample that also contains observations on managers entering into the firm at levels higher than Level 1. In that version I allow for h_1 to vary across managers entering at different levels and this variation turns out to be quantitatively important for job assignment. (The estimates of the key parameters across the two samples, however, are very similar if not indistinguishable. See the Supplementary Appendix for details.)

Lastly, observe that this specification of the human capital process implies that acquired human capital does not affect belief updating. Nonetheless, as is apparent from the term $\alpha_{Ak}A_{kt}$ in (21), this formulation allows for an interaction between unobserved initial ability, as captured by $\theta \in \{\alpha, \beta\}$, and observed acquired technological human capital, as captured by h_t, which can lead to the expected output of managers of different ability to change over time at different rates. For instance, if $\alpha_{Ak} > \beta_{Ak}$ and $A_{kt} > 0$, then $\alpha_{Ak}A_{kt} > \beta_{Ak}A_{kt}$, so the same skills acquired by managers of high (α) and low (β) ability, as measured by A_{kt}, imply a higher expected output for a manager of high ability than for one of low ability.

This feature of the human capital process is modeled in the spirit of BGH. In addition to providing convincing descriptive evidence of the importance of learning about ability and the presence of selection on unobserved productive characteristics in their data, BGH (p. 903) conclude that their data are consistent with an interpretation of unobserved ability in an expansive way, not only as an unobserved productivity parameter but also as the unobserved rate at which managers accumulate new skills.

Unobserved Heterogeneity. I also allow managers to differ in ways that are unobserved by the econometrician but observed by both managers and firms. Allowing for additional dimensions of unobserved heterogeneity among managers addresses potential concerns about sample selection that may arise in my framework for two reasons. First, neither proxies for managers' ability endowments, other than education and age, nor information about managers' labor market histories before entry into my firm are observable in my data. Yet managers may self-select in my sample, by entering into the firm, based on these unobserved characteristics and outcomes. Second, my data do not contain information about managers' careers once they leave the firm.

I correct for the first potential source of selection by allowing the initial prior about a manager's ability to vary unobservably across managers and by allowing for heterogeneity in (some of) the parameters of the distribution of wages at my firm. This heterogeneity in wages is meant to capture the fact that persistent productivity differences among managers may be observable to outside firms through resumes or job interviews. If such differences affect a manager's performance at these firms, they are then reflected in the wages paid my firm, due to competition between my firm and these alternative prospective employers. I address the second potential source of selection both by modeling the reasons for endogenous separations and by allowing for the possibility of exogenous separations.

Formally, I assume that managers can be one of I skill types, indexed by $i = 1, \ldots, I$, in the spirit of Heckman and Singer (1984). See Keane and Wolpin (1997) and Eckstein and Wolpin (1999) for an illustration of the usefulness of this nonparametric specification of unmeasured persistent traits of individuals in the context of dynamic nonlinear structural estimation frameworks. I allow types to differ along two dimensions. First, different manager skill types can differ in their initial priors. Specifically, I assume that the initial prior that a manager is of high ability can take on values p_{i1} with probability q_i for $i = 1, \ldots, I$. In light of changes in likelihood values, evaluated based on a penalized likelihood criterion (the Akaike information criterion) and quantitative implications of the model, I set I equal to four. Doing so yields a total of seven additional parameters, corresponding to the points in the support of the type distribution and its mass points, $(\{p_{i1}\}_{i=1}^{4}, q_1, q_2, q_3)$, which are estimated together with the rest of the model's parameters. Second, I allow different manager skill types to differ in the parameters governing wages paid at my firm, as I explain later.

Allowing for unobserved permanent characteristics that affect priors and wage parameters renders the econometric model a semiparametric mixture. Clearly, such a mixture model is more flexible than the parametric model originally specified. In addition, it helps me isolate in a parsimonious way the importance of self-selection based on productivity characteristics known to agents (but unobservable to the econometrician) from the importance of learning about characteristics unknown to agents (and to the econometrician) for the career patterns of the managers in my sample. (For evidence of self-selection on characteristics unobservable to the econometrician in another study of the careers of professionals, see Sauer (1998) for an analysis of the careers of lawyers.)

4.2 Job Assignment and Separation

I turn now to deriving the probabilities of job assignment and separation. I consider the event of a separation between the firm and a manager as corresponding to the choice of Level 0. For simplicity I now drop the firm notation.

In the presence of unobserved heterogeneity, the beginning-of-period state for a manager of skill type i is $s_{it} = (p_t, h_t, i)$ where $s_{i1} = (p_1, h_1, i)$ at $t = 1$ and, by the process of technological human capital acquisition, $s_{it} = (p_t, h_1, t-1, k_{t-1}, i)$ at $t \geq 2$. As discussed, observed human capital at entry into the firm, h_1, as captured by a manager's entry age, education, and year of entry, did not prove empirically to affect the probability of observed assignment or separation in any significant way, conditional on the other state variables. So, I omit the dependence on h_1 here. As for the impact of i on job assignment and separation, note that all parameters of productivity and technological human capital acquisition may in principle depend on i. From the point of view of job assignment and separation, however, only the dependence of beliefs on i has turned out to be important in estimation. Then, the portion of the beginning-of-period state sufficient for the purpose of retention and job assignment reduces to $(p_{it}, t-1, k_{t-1})$.

By (18) I can express the match surplus value function between my firm and a manager as

$$V(p_{it}, t-1, k_{t-1}, \boldsymbol{\varepsilon}_t) = \max_{k \in K \cup \{0\}} \{(1-\delta)[y(p_{it}, t-1, k_{t-1}, k) + \varepsilon_{kt}]$$

$$+\delta(1-\eta_{kt}) \int_{\boldsymbol{\varepsilon}_{t+1}} EV(p_{it+1}, t, k, \boldsymbol{\varepsilon}_{t+1} | p_{it}, t-1, k_{t-1}, k) dF(\boldsymbol{\varepsilon}_{t+1}) \Bigg\}, \quad (23)$$

at $t \geq 2$, where $y_A(s_{it}, k)$ specializes to $y(p_{it}, t-1, k_{t-1}, k)$; ε_{Akt} and η_{Akt} specialize to ε_{kt} and η_{kt}; and $y_f(s_t, k_{ft})$, $\varepsilon_{fk_{ft}t}$, and $\eta_{fk_{ft}t}$ specialize to $y(p_{it}, t-1, k_{t-1}, 0)$, ε_{0t}, and η_{0t}. The expression for $V(p_{i1}, \boldsymbol{\varepsilon}_t)$ at $t=1$ can be obtained analogously, as shown in the Supplementary Appendix. As I explain in the Supplementary Appendix, the retention and job assignment problem in the eighth year of tenure of a manager reduces to an infinite-horizon match surplus maximization problem. The solution to this problem endogenously provides the terminal value for the backward induction routine that I use to solve for the firm's retention and job assignment problem between the first and the seventh year of tenure of a manager, the main years of interest.

I assume that productivity shocks have multivariate type I extreme value distributions, each with mean zero and (normalized) variance $\pi^2/6$. Under this assumption, by applying the results of Rust (1987, 1994) to (23) it follows that at any tenure between $t=2$ and $t=7$,

$$V(p_{it}, t-1, k_{t-1}, \boldsymbol{\varepsilon}_t) = \max_{k \in \{0,1,2,3\}} \{v(p_{it}, t-1, k_{t-1}, k) + \varepsilon_{kt}\},$$

where the alternative-specific match surplus value $v(p_{it}, t-1, k_{t-1}, k)$, $1 \leq k \leq 3$, equals

$$(1-\delta)y(p_{it}, t-1, k_{t-1}, k) + \delta(1-\eta_{kt})r_k(p_{it}) \log\left\{\sum\nolimits_{k' \in \{0,1,2,3\}} \exp\left[v(P_{Hk}(p_{it}), t, k, k')\right]\right\}$$

$$+\delta(1-\eta_{kt})[1-r_k(p_{it})] \log\left\{\sum\nolimits_{k' \in \{0,1,2,3\}} \exp\left[v(P_{Lk}(p_{it}), t, k, k')\right]\right\} \quad (24)$$

and $r_k(p_{it}) = \alpha_k p_{it} + \beta_k(1-p_{it})$. Similar expressions can be obtained for $t=1$ and $t=8$; see the Supplementary Appendix for details.

Recall that in the model managers separate for both exogenous and endogenous reasons. An endogenous separation occurs when a competing firm, namely, the second-best firm, makes a manager at my firm a more attractive offer (and my firm is indifferent between employing and not employing the manager). When this happens, the alternative $k=0$ provides the largest match surplus value. The model implies that the match surplus value from choosing $k=0$ is a continuous function of the prior. I specify this continuous function, $v(p_{it}, t-1, k_{t-1}, 0)$, as a flexible υ-th order polynomial in p_{it},

$$v(p_{it}, t-1, k_{t-1}, 0) = \nu_{0t} + \cdots + \nu_{\upsilon t} p_{it}^{\upsilon} \tag{25}$$

with $\upsilon \geq 1$. In (25) the dependence of the parameters on tenure is to account for the effect of h_t on the match surplus value of separation. See Erdem and Keane (1996) and Ching (2010) for similar parameterizations of the value of the 'reference alternative' in learning models of dynamic discrete choice. By applying the results of Rust (1987, 1994), the probabilities of assignment to Level k, $0 \leq k \leq 3$, in each tenure year can then be straightforwardly computed, analogously to a standard multinomial logit model. See the Supplementary Appendix for all omitted details.

Note that the match surplus maximization problem, which is a (nonlinear semiparametric mixture) dynamic discrete choice problem, entails a repeated choice among $K+1$ alternatives whose values are functions of p_{it}. Naturally, by treating the second-best firm as the reference alternative for this problem, at most terms of degree higher than one in (25) can be identified. Hence, I interpret all estimated parameters of productivity and (technological) human capital accumulation, a_{kt}, $b_{kk_{t-1}t}$, and c_{kt}, as differences with respect to the corresponding parameters of the second-best firm, ν_{0t} and ν_{1t}, which I set at zero. In practice, for all relevant sets of parameters, the coefficients on prior terms of degree higher than one in (25) have proved insignificant. The values $v(p_{it}, t-1, k_{t-1}, k)$, $1 \leq k \leq 3$, have proved to be approximately linear in p_{it} too.

4.3 Recorded Performance Ratings

Here I describe how I specify the error structure that relates the observed distribution of recorded performance ratings of managers at a level to the distribution of output realizations implied by the model. Note first that my data directly record the performance ratings of each manager at the firm, even when a manager is engaged in a group project. Then, analogously to BGH, I interpret the performance rating of a manager as reflecting that manager's contribution to firm value in a particular period. So, if actual performance signals are complementary, for instance, across managers who work together, I assume that this complementarity is taken into account by supervisors when expressing their evaluation of a manager's performance through the recorded rating.

A potentially important source of error in recorded ratings is the tendency of supervisors to assign uniform ratings to employees regardless of their actual performance. This tendency is well known and has been discussed by Baker, Jensen, and Murphy (1988), Murphy (1992), and Prendergast (1999), among others. This type of error may lead to systematic misreporting of actual performance and, thus, to bias. Since performance ratings play a critical role in my model, by driving the dynamics of belief updating, I pay special attention to flexibly modeling error in recorded performance. In addition, if misclassification is present in the data but not explicitly accounted for, then maximum likelihood estimation may lead to biased and inconsistent parameter estimates (see Hausman, Abrevaya, and Scott-Morton (1998)).

Formally, based on Keane and Wolpin (1997) and Keane and Sauer (2009), I model error in recorded performance ratings as follows. Let R^o_{nt} denote manager n's observed performance in period t and R_{nt} denote the manager's true performance realized in period t. Let L^o_{nt} denote the manager's observed job assignment. Since I maintain that the error in observed ratings is independent of the type of a manager, no variable is indexed by i.

I specify the error in recorded performance ratings as follows. Let $E_0(L^o_{nt}, t) = \Pr(R^o_{nt} = 1 | R_{nt} = 0, L^o_{nt}, t)$ denote the probability of a recorded high rating when a low rating is realized in job L^o_{nt} at tenure t and let $E_1(L^o_{nt}, t) = \Pr(R^o_{nt} = 0 | R_{nt} = 1, L^o_{nt}, t)$ denote the probability of a recorded low rating when a high rating is realized in job L^o_{nt} at tenure t. I assume that

$$E_0(L^o_{nt}, t) = \frac{\exp\{d_0 + d_2(L^o_{nt})[tI(L^o_{nt} = 1) + (t-1)I(L^o_{nt} = 2)]\}}{1 + \exp\{d_0 + d_2(L^o_{nt})[tI(L^o_{nt} = 1) + (t-1)I(L^o_{nt} = 2)]\}},$$

$$E_1(L^o_{nt}, t) = \frac{1}{1 + \exp\{d_0 + d_1 + d_2(L^o_{nt})[tI(L^o_{nt} = 1) + (t-1)I(L^o_{nt} = 2)]\}}.$$

First, note that I let these errors vary both with a manager's assigned job and with the manager's tenure at my firm. That the errors may vary with the assigned job seems natural, since the evaluation of performance is often job-specific. Letting the errors vary with tenure is one way to capture the possibility that performance appraisal may be conducted more thoroughly at certain stages in a manager's career at the firm, such as when managers are newly hired or become eligible for new tasks. Observe that in this specification, the greater is $d_2(L^o_{nt})$, the greater is the persistence in misreporting. Since no individual is observed at Level 2 at entry, I let the errors at Level 2 depend on $t-1$ rather than t.

Second, note that the classification error rates, $E_0(\cdot)$ and $E_1(\cdot)$, are directly informative as to whether the firm uses realized or recorded performance in its assignment decisions. Indeed, as $E_0(\cdot)$ and $E_1(\cdot)$ approach zero, the probability of a correct classification approaches one, so recorded performance coincides with realized performance.

Third, note that, since the probability of a recorded rating is not a linear function of its true probability, such a classification error scheme leads to bias (Keane and Sauer (2009)).

To limit parameter proliferation, and based on model diagnostic and fit, I maintain that $d_1 = d_0$. Also, I estimate classification error parameters only at Levels 1 and 2 because among managers assigned to Level 3, many ratings are missing, whereas few are missing at Levels 1 and 2. Specifically, ratings are missing for no fewer than 35 percent of individuals assigned to (the original) Level 3 and higher, with much higher proportions of missing values, up to 44.9 percent, at lower tenures. Moreover, the distribution of ratings at Level 3 and higher in the data differs from the distribution at lower levels. Estimation of classification error rates at the higher levels thus proved problematic. As a result, the estimated parameters governing the distribution of performance ratings are $(\{\alpha_k, \beta_k\}_{k=1}^3, d_0, d_2(1), d_2(2))$.

4.4 Wages

Here I discuss my empirical specification of the wages paid by my firm to its managers. Note that in expressions (9) and (10) I have notation for a single manager only. In estimation I allow for both unobserved heterogeneity, indexed by i, and for observed heterogeneity, indexed by $n = 1, \ldots, N$, among managers; so now I index a manager by in. In terms of observed characteristics, in the data I observe managers' age at entry into the firm, their number of years of completed education at entry, and their calendar year of entry. I record these observed characteristics as the vector $e_n = (age_n, edu_n, year_n)$. Note that these three variables refer to managers' characteristics at entry into the firm and, hence, do not vary over time or with tenure.

By (1), as mentioned I interpret the characteristics in e_n as capturing the component of a manager's human capital at entry into the firm, denoted now by h_{n1}, that is transferable to other firms and thus, due to competition, affects wages at my firm. Formally, let h_{n1} be some invertible function of $e_n = (age_n, edu_n, year_n)$, so that from h_{n1} I can recover each component of e_n, and let $h_{nt} = (h_{n1}, t-1, k_{t-1})$. So, manager in has state $s_{in1} = (p_1, h_{n1}, i)$ at $t = 1$ and $s_{int} = (p_t, h_{n1}, t-1, k_{t-1}, i)$ at $t \geq 2$. To make apparent the dependence of beliefs on a manager's skill type, I denote the prior at t that manager in is of high ability by p_{it}. It turns out that conditional on the prior, p_{it}, human capital at entry, h_{n1}, and tenure, $t-1$, wages at each level effectively do not depend on previous job assignment, k_{t-1}. Hence, empirically, the prior updated over time through Bayes' rule already summarizes the impact of k_{t-1} on wages. For this reason, no explicit dependence on k_{t-1} appears in the expressions for paid wages.

Consider now the term in (9) related to a manager's expected output at the second-best firm f. The model implies that the identity of that firm is a function of s_{int} and $\boldsymbol{\varepsilon}_t$. Recall that the expected output of manager in at the second-best firm (after productivity shocks are realized) is given by $y_f(s_{int}, k_{ft}) + \varepsilon_{fk_{ft}t}$. I then express the expected output at the second-best firm of manager in at tenure t at my firm as

$$y_f(s_{int}, k_{ft}) + \varepsilon_{fk_{ft}t} = \varpi_{inkt} + \bar{\omega}_{ik} p_{it} + \varepsilon_{inkt}. \tag{26}$$

Note that even though the identity of the second-best firm depends on $\boldsymbol{\varepsilon}_t$ through $f = f(s_{int}, \boldsymbol{\varepsilon}_t)$, this dependence is irrelevant. The reason follows from the discussion after Proposition 3: given my assumption that all other firms have replicas, my economy is equivalent to one in which a single outside firm has access to all the technologies of the replica firms and sets wages competitively. Therefore, none of the coefficients of $y_f(s_{int}, k_{ft})$ depends on $\boldsymbol{\varepsilon}_t$ directly and, given the assumption that the shocks are type I extreme value, the additive term ε_{inkt} on the right-hand side of (26) captures exactly the impact of $\boldsymbol{\varepsilon}_t$ on the expected output of the second-best firm. The dependence of the terms on the left-side of (26) on f is thus subsumed by the dependence of the terms on the right-side of (26) on p_{it}, h_{nt}, i, and k. (The dependence of ε_{inkt} on i and k is through its variance, as discussed momentarily; the dependence of $\bar{\omega}_{ik}$ on n and t is suppressed as irrelevant in practice.) Hence, $\varpi_{inkt} + \bar{\omega}_{ik}p_{int}$ denotes manager in's ability-dependent component of expected output at firm f and the error term ε_{inkt} is short-hand for $\varepsilon_{fk_{ft}t}$, the productivity shock realized at the job k_{ft} that firm f offers to manager in assigned to job k at my firm at tenure t. I specify ϖ_{inkt} as

$$\varpi_{inkt} = \varpi_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1),$$

where the intercept term $\varpi_{ik}(age_n, edu_n, year_n)$ is given by

$$\omega_{0ik} + \varpi_{1k}age_n + \varpi_{2k}age_n^2 + \varpi_{3k}edu_n + \sum\nolimits_{m=1}^{9} \varpi_{ym} I(year_n = m). \tag{27}$$

As implied by the model, in (26) the prior of a manager enters in a linear fashion. Tenure at the firm, an element of h_{nt}, is weighted by the coefficient ω_{1kt}. Finally, the initial human capital h_{n1}, the only other relevant element of h_{nt}, is captured by the vector $(age_n, edu_n, year_n)$, which represents the tenure- and time-invariant component of ϖ_{inkt}. More precisely, in (27) the coefficients ϖ_{1k}, ϖ_{2k}, ϖ_{3k}, and ϖ_{ym} capture the degree of transferability to other firms of the technological human capital acquired by managers before entry into my firm. In particular, $I(year_n = m)$ is an indicator function that equals one if, and only if, the year of entry of the manager is m, where $m = 1$ corresponds to 1971, $m = 2$ corresponds to 1972, and so on. These coefficients ϖ_{ym} allow for an interaction between h_{n1} and the calendar year of entry of a manager into the firm in order to capture the possibility that business cycle conditions at the time of entry into the firm may affect the value of h_{n1} to firms in the market and, hence, paid wages at my firm. Note that Baker, Gibbs, and Holmström (1994a,b) find that such cohort effects are important, even after several years of tenure; see also Waldman (forthcoming) on this point.

Next, consider the term Ψ_{ift} for the wage premium that, as discussed, my firm pays to compensate a manager for the lost information and human capital that the manager could have acquired by working at the second-best firm in period t. As consistent with the model by

(10), I flexibly specify Ψ_{ift} as an r-th order polynomial in the prior p_{it},

$$\Psi_{ift} = \psi_{ik0} + \ldots + \psi_{ikr} p_{it}^r \tag{28}$$

with $r \geq 1$. (In (28) I only record the dependence on the state that has proved to be empirically relevant.) Note that the dependence of the learning premium Ψ_{ift}, as well as of $\varpi_{ik}(\cdot)$ and ω_{1kt}, on level assignment at my firm is to capture the possibility that different jobs at my firm offer different prospects of acquisition of technological human capital, which may be transferable to other firms.

I now discuss the error structure for wages. Clearly, unmeasured aspects of the firm's wage policy or of a manager's behavior, as well as recording error, may influence paid wages. I allow these unmeasured characteristics and recording error to idiosyncratically affect a manager's wage. I denote this additional source of stochastic variability by ε_{inkt}^m, which I assume is a generalized extreme value random disturbance, and denote by u_{inkt} the sum of the productivity shock at the second-best firm, ε_{inkt}, and ε_{inkt}^m, so

$$u_{inkt} = \varepsilon_{inkt} + \varepsilon_{inkt}^m. \tag{29}$$

Since the random variable u_{inkt} is the sum of a type I extreme value shock, ε_{inkt}, and a generalized extreme value shock, ε_{inkt}^m, it is logistically distributed, so it can be well approximated by a normal random variable.[10] Hence, I treat u_{inkt}, $1 \leq k \leq 3$, as normally distributed with mean zero and variance σ_{ik}^2 depending on a manager's skill type and current job assignment at my firm. The dependence of the variance of u_{inkt} on a manager's skill type accounts for the possibility that managers of different skill types have access to different labor market opportunities and, hence, different prospective employers. I also allow the variance of u_{inkt} to depend on a manager's assignment in order to capture the possibility that a manager's task or position within the firm's hierarchy allows the manager to come into contact with different firms. This possibility as well may induce idiosyncratic variation in the identity of the firm offering a manager the second-highest value of wages.

Since managers are salaried employees, it is convenient to specify a wage per unit of labor input. The idea is that if managers are paid proportionally to their labor input in a period, λ_{int}, then $w_{int}^o = w_{int}\lambda_{int}$, where w_{int} is the wage implied by the model. Thus, I interpret the model as determining a process for $\ln(w_{int}^o)$. By (9), (10), and the above assumptions, I specify the log of the observed wage of manager in at tenure t as the sum of (26), $\delta\Psi_{ift}/(1-\delta)$ from

[10] Recall that if a random variable X has a Gumbel distribution with parameters α and β and another random variable Y has a generalized extreme value distribution with parameters α, β, and 0, then the sum of X and Y is a random variable logistically distributed with parameters 2α and β.

(28), and ε^m_{inkt}, which can be expressed as

$$\ln(w^o_{int}) = \varpi_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1) + \bar{\omega}_{ik}p_{it} + \delta(\psi_{ik0} + \ldots + \psi_{ikr}p^r_{it})/(1-\delta) + u_{inkt}.$$

Define $\varpi_{0ik} = \omega_{0ik} + \delta\psi_{ik0}/(1-\delta)$, $\omega_{2ik} = \bar{\omega}_{ik} + \delta\psi_{ik1}/(1-\delta)$, and $\omega_{3i} = \delta\psi_{ik2}/(1-\delta)$. Then,

$$\ln(w^o_{int}) = \omega_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1) + \omega_{2ik}p_{it} + \omega_{3i}p^2_{it} + u_{inkt},$$

$$\omega_{ik}(age_n, edu_n, year_n) = \varpi_{0ik} + \varpi_{1k}age_n + \varpi_{2k}age^2_n + \varpi_{3k}edu_n + \sum\nolimits_{m=1}^{9} \varpi_{ym}I(year_n = m) \quad (30)$$

follow by collecting and rearranging terms. I have ignored terms in Ψ_{ift} of degree higher than two since they have proved insignificant for all relevant sets of parameters. Also, since ω_{2ik} has proved not to differ substantially across levels conditional on i, I restrict $\omega_{2ik} = \omega_{2i}$. Lastly, given that ω_{3i} has turned out not to vary across different manager skill types in any significant way, I restrict $\omega_{3i} = \omega_3$. Econometrically, observe that the dependence of some terms of the wage equation on i implies that the estimated wage equation is a semiparametric mixture of parametric components.

Finally, in contrast to wage specifications common in the applied literature, the time path of wages (apart from the tenure trend in the intercept term in (30)) is solely due to the dynamics of belief updating, endogenous human capital acquisition, and competition. In practice, the tenure coefficient ω_{1kt} has proved significantly different from zero only at Level 1; see the later discussion of the parameter estimates.

4.5 Identification

Here I provide an intuitive overview of identification in my model, relegating a more formal discussion to the Supplementary Appendix. To investigate the identifiability of the model parameters in practice, I also performed a Monte Carlo exercise, reported in the Supplementary Appendix, which provides evidence in support of the model being identified. Specifically, the results of the Monte Carlo analysis show that parameters are estimated with small bias and that asymptotic standard errors closely track the empirical standard deviations of the parameters.

Broadly, the focus of the learning component of my model is the acquisition of *informational* human capital by managers, which manifests itself in changes in the prior over time. Several features of the data about careers within firms point to the importance of learning. First, job mobility in firms, which is large, is closely linked to wage growth. (See Lazear (1992), Gibbons and Waldman (1999a, 1999b), Gibbs and Hendricks (2004), Belzil and Bognanno (2008), and Waldman (forthcoming) for similar evidence.) Second, wages are greatly dispersed among workers at any given level and their variance increases over time. Third, demotions from higher- to lower-level jobs are rare, but decreases in real wages are common. (See Baker, Gibbs, and Holmström (1994a) and the survey by Gibbons and Waldman (1999a).)

As discussed earlier, in the model I also allow for the acquisition of *technological* human capital by managers, resulting from the accumulation of new productive skills through employment. Here I first explain how I can separately identify these two types of human capital. I then explain how I can identify the degree of transferability of technological human capital across levels within my firm and, finally, across firms.

Informational vs. Technological Human Capital. To understand at an intuitive level how I can distinguish between informational and technological human capital, it is useful to compare the predictions of the model when only one of these two types of human capital is present.

If I allow for only the accumulation of informational human capital, then whenever high ability is more valuable at high-level jobs, this pure learning version of my model typically predicts a positive fraction of demotions from higher- to lower-level jobs. The reason is that a long enough sequence of low performance ratings causes the prior to decrease to the point at which demotion to a lower-level job is optimal. In the data, however, I observe no demotions. When I include technological human capital as well, the acquisition of new skills can make a manager's value to the firm grow fast enough to offset the downward revision in the prior caused by poor performance. Thus, the model with both types of human capital can predict promotions and the associated wage increases, wage decreases, and no demotions. More precisely, the zero frequency of demotions at each level and tenure constitutes a set of moments that lend support to the hypothesis that some form of acquisition of new skills is present in my data.

In contrast, if I allow for only technological human capital (with no depreciation or productivity shocks), then this pure skill acquisition version of my model predicts no real wage decreases for managers continuously employed at my firm. In the data, however, in every year, more than 20 percent of wage changes are negative (Table 5). Indeed, Gibbons and Waldman (1999a,b) argue that negative real wage changes are an important feature of the dynamics of wages in firms, consistent with the idea that firms and workers learn about the value of their matches. Indeed, when I include in the model informational human capital as well, low performance leads to lower priors and, hence, naturally to real wage decreases. Thus, the sizeable fraction of negative real wage changes at each level and tenure constitutes a set of moments that lend support to the hypothesis that some form of learning is present in my data.

Transferability of Technological Human Capital. To understand how I can identify the transferability of human capital across levels at my firm, observe first that the pure learning version of my model predicts that, given the prior, the probability of promotion is independent of tenure. If I add technological human capital that is nontransferable across levels, that is, purely task-specific, then the probability of promotion decreases with tenure at a level, conditional on the prior. The reason is that greater task-specific human capital makes a manager better suited to the level the manager is assigned to. If, instead, the technological human capital acquired with tenure at a given level substantially improves performance at higher levels, then, given the prior, the probability of promotion increases with tenure at the level. As a consequence, the tenure profile of the hazard rate of promotion at each prior and level constitutes a set of moments that help identify the degree of transferability of human capital across levels.

Finally, to understand how I can identify the transferability of human capital across firms, observe that if acquired human capital were purely firm-specific, then, conditional on the prior, wages on average would be independent of a manager's tenure or job assignment at my firm. In the data, even after conditioning on a manager's prior and entry characteristics, paid wages depend also on tenure and job assignment. Therefore, statistics on the distribution of wages by level and tenure constitute a set of moments that help identify the degree of transferability of human capital across firms.

5 Estimation Results

In this section, I first present evidence about the fit of the model to the data and then discuss some of the key parameter estimates, commenting on their implications for the sources of observed job and wage mobility. In the Supplementary Appendix I derive the likelihood function and discuss in detail the numerical solution of the model. I also show that results of a Monte Carlo exercise support the identifiability of the model's parameters in practice. There I also explain the inferences that, based on these parameter estimates, can be drawn about the technological characteristics of other firms in the market that compete for the managers employed at my firm.

Here I confine attention to estimates obtained from the sample of managers entering into the firm at Level 1. To address potential concerns about selection that such a sample choice may raise, I have reestimated the model on a sample that also includes observations on managers entering the firm at higher levels. I found estimates very similar to those obtained from the sample of entrants at Level 1, and the model fit improved. Specifically, relative to the results reported here, the parameters are more precisely estimated. For the two specifications I estimate on this larger sample, distinguished by the flexibility I allow for the error structure for wages at Level 3, the fit of the model is almost perfect in terms of the distributions of level assignments and performance ratings, and in terms of the wage distributions for the second, more flexible, specification. Also, based on the Pearson's χ^2 test, the model is almost never rejected. I interpret these results as evidence that the model is quite successful at capturing the patterns of interest in the data and that the estimation results presented here are robust. For details, see the Supplementary Appendix.

5.1 Model Fit

Here I present evidence on the ability of the estimated model to capture the main patterns of interest in the data. Specifically, I evaluate the fit of the model by comparing observed and predicted outcomes along three dimensions: (1) the distribution of managers across levels by tenure and the hazard rates of separation, retention at a level, and promotion to the next level at each level by tenure; (2) the distribution of performance ratings at Levels 1 and 2 by tenure; and (3) the distribution of wages at each level by tenure. I relegate the goodness-of-fit test results to the Appendix. In assessing model fit (and later conducting the counterfactual experiments), I simulated 4,000 prior realizations per manager, drawn from the estimated nonparametric distribution of initial priors.

Overall, as Tables 6–9 make clear, the model successfully captures the tenure profile of separation and assignment to Levels 1, 2, and 3, as well as the distribution of performance ratings at Levels 1 and 2 at each tenure. The model also fits quite well the wage distribution at each level and tenure. I turn now to discuss in more detail how the model fits the data along the three dimensions considered.

Level Assignments. Consider the distribution of managers across levels and the associated hazard rates of separation and promotion.

Table 6 compares the distribution of managers across levels observed in the data and those predicted by the model by year of tenure. Notice that the model tracks remarkably well, both qualitatively and quantitatively, the tenure profile of assignment to Levels 1, 2, and 3 and the pattern of manager separation. The predicted percentage of managers at Level 1 rapidly decreases with tenure at the firm as in the data, whereas the predicted percentage assigned to Levels 2 and 3 first increases, until the third and sixth years of tenure, respectively, and then decreases, just as in the data. In addition, both the predicted and actual percentages of managers who separate from the firm are substantial in each year.

Table 7 presents the hazard rates of separation, retention at a level, and promotion at each level by tenure for the data and the model. Note that the model accurately reproduces the feature that outflows from Level 1 come from an essentially constant hazard rate of separation and an increasing, then decreasing, hazard of promotion. Correspondingly, the inflow of managers into Level 2 comes almost exclusively from promotions from Level 1, which at first increase, then decrease, with tenure. Indeed, the predicted fraction of demotions is negligible. Note also that the model predicts, as in the data, that outflows of managers from Level 2 come from an essentially constant hazard rate of separation and a hazard rate of promotion to Level 3 that first increases, then decreases, with tenure. With respect to Level 3, in the model as in the data, the inflows of managers come overwhelmingly from promotions from Level 2, which at first increase and then decrease with tenure. In fact, the predicted fraction of managers promoted from Level 1 to Level 3 is negligible. As in the data, the outflows from Level 3 come overwhelmingly from separations, which, as at the other levels, are characterized by an essentially constant hazard rate.

The model's quantitative implications for these hazards are also very much in line with the data. The largest discrepancies between the model and the data are mainly at Level 1 for the hazard rate of retention between the second and third years of tenure and for the hazard rate of promotion between the third and fourth years of tenure, and at Level 2 for the hazard rate of promotion to Level 3 between the second and third years of tenure. Note, however, that small differences in the predictions for the level assignments of managers (in Table 6) can lead to relatively large differences in the predictions for hazard rates, especially since over time hazard rates are computed on cells of observations of smaller and smaller size. Indeed, quite surprisingly, the model tracks quite well the hazard rate of retention at Level 1 and promotion to Level 2 at high tenures, even though few managers are still assigned to Level 1. Overall, the model successfully captures the qualitative and quantitative characteristics of the tenure profile of the hazard rate of retention at Levels 1, 2, and 3, and promotion to Levels 2 and 3.

Performance Ratings. Table 8 displays the distribution of performance ratings at Levels 1 and 2 by tenure for the data and the model. Note that the model successfully fits the patterns in the data. Clearly, the model matches well the fact that the percentage of high ratings at each level decreases with tenure. It also matches the fact that at any given tenure the percentage of high ratings increases with the level. In fact, the model tracks remarkably well the percentage of high ratings at almost all tenures at Level 1 (with slight overpredictions in the third, fifth, and seventh years of tenure) and at Level 2 (apart from some discrepancies in the fourth and seventh years of tenure).

Wages. The distribution of wages by level and tenure are displayed in Table 9 for the data and the model. Clearly, the model fits well these distributions by level and tenure, apart from slight discrepancies at Level 3 at the highest tenures. In particular, the model reproduces the fact that the mean wage, across tenures, increases with the level, as well as the fact that the mean wage decreases with tenure at Leve1 1 and is approximately constant at the higher levels. It also matches the fact that the standard deviation of wages is higher at higher levels. (See Table 11A.) The largest discrepancy between the observed and predicted distribution of wages is at Level 3 in the sixth and seventh years of tenure, but this difference is partly due to the large rate of attrition in the sample. Recall that almost 60 percent of managers have left the firm by the seventh year after entry. Finally, as in the data, the model implies that the yearly fraction of wage decreases in each tenure year is no less than 20 percent.

5.2 Parameter Estimates and Implications

Here I discuss the estimates of the main parameters of interest, namely, those related to initial uncertainty, learning, error in recorded performance ratings, productivity and technological human capital acquisition, turnover, and wages. I also discuss their implications for the degree of market power of my firm and for wage growth on the job. All parameter estimates and their asymptotic standard errors, computed based on the outer product of the scores of the loglikelihood function, are displayed in Table 10. Note that all parameters prove significant at the 1 percent level.

Uncertainty, Learning, and Ratings Error. Observe that the parameters for initial uncertainty capture the scope for learning through employment, the parameters for learning at each job capture both the nature of the experimentation problem that my firm faces as well as the speed at which learning occurs, whereas the parameters for errors in performance capture the extent to which this learning process is contaminated by error in the data.

Here is a summary of my results. I find, first, that initial uncertainty is large and heterogeneous among managers' skill types, in that initial priors are close to uninformative but dispersed across different types of managers. Second, I find that since the ordering of jobs by the success rate of a manager of high or low ability is the opposite of their ordering by the informativeness of performance, my firm faces a nontrivial multi-armed bandit problem when assigning workers to jobs. Third, I find that learning takes place slowly over time. Fourth, I find that the estimated heterogeneity in initial priors interacts with the parameters governing success and failure at each job so as to imply substantial differences in the speed of learning across managers of different skill types. Fifth, I find that the error in recorded ratings displays an interesting pattern: the probability of an incorrect high rating decreases with tenure, whereas the probability of an incorrect low rating increases with tenure.

I first show that the parameter estimates imply that the degree of uncertainty about a manager's ability is large at entry into the firm. Recall that initial uncertainty about ability is captured by the distribution of the unobserved skill types of managers, which is also the distribution of possible initial priors. Formally, each type i is characterized by a prior probability that the manager's ability is high, p_{i1}, and by the fraction of managers of this type, q_i. (Each type is also characterized by certain parameters of the distribution of wages; see later discussion.) I estimate that 15.5 percent of managers are of the first skill type and have a prior of 0.338; 21.1 percent are of the second skill type and have a prior of 0.381; 31.3 percent are of the third skill type and have a prior of 0.465; and 32.1 percent are of the fourth skill type and have a prior of 0.607. (Note that to avoid boundary problems in estimation, I express $p_{i1} = \exp\{\phi_{i1}\}/[1 + \exp\{\phi_{i1}\}]$ and estimate ϕ_{i1}, a parameter that ranges over the real line, instead of p_{i1}. In Table 10 I report parameter estimates and standard errors for ϕ_{i1}.) Clearly, these estimates imply a large degree of uncertainty about a manager's ability: the average prior probability that a newly hired manager is of high ability is close to 0.50 (specifically, $\sum_i^I q_i p_{i1} = 0.473$). Initial beliefs also display a high degree of variability across managers of different skill types: a manager of the fourth skill type is nearly twice as likely to be of high ability than a manager of the first skill type. Taken together, the finding that priors are mostly intermediate and their heterogeneity among managers of different skill types point to a potentially important role for learning in explaining the variability of the employment experience of managers with similar observed characteristics at entry.

Second, I contrast the ordering of jobs by the informativeness of performance with their ordering by the success rate of managers of high and low ability. Learning is determined by the values of α_k and β_k that govern the Bayesian updating of beliefs at each Level k. At Level 1 they are $(\alpha_1, \beta_1) = (0.514, 0.456)$; at Level 2, $(\alpha_2, \beta_2) = (0.5437, 0.491)$; and at Level 3, $(\alpha_3, \beta_3) = (0.5435, 0.490)$. Hence, managers of either high or low ability are most likely to succeed at Level 2, second most likely to do so at Level 3, and least likely to do so at Level 1. These estimates also imply that I can order jobs by the informativeness of performance using the *Blackwell criterion* (Blackwell (1951)). This criterion orders the informativeness of experiments—here, the assignment of a manager to a job—according to the second-order stochastic dominance ordering of the posterior beliefs reached after observing the experiment's outcome—here, the manager's performance. According to that criterion, the most informative job is Level 1, the second most informative is Level 3, and the least informative is Level 2.[11] Thus, the ordering of job levels by the likelihood of success of managers of high and low ability, $(2, 3, 1)$, is the opposite of their ordering by the informativeness of performance, $(1, 3, 2)$. These patterns imply that the firm in my data faces a trade-off between assigning managers to jobs according to the probability of success of a manager and assigning managers to jobs according to the informativeness of performance. Hence, the firm indeed faces an interesting multi-armed bandit problem.

Third, I explain how the estimates for α_k and β_k imply that learning is a slow process. To see this, note that it takes 20 years of consecutive high performance at Level 1 for the average prior about a manager's ability being high, 0.473, to increase to 0.90. Starting with the same prior at Level 2 or 3, this process takes 23 years. This finding is consistent with the fact that uncertainty at entry into the firm is large even for managers who arguably have several years of experience in the labor market. (This finding is also mirrored by the results on the speed of learning at the second-best firm. See the Supplementary Appendix.)

Fourth, the speed of learning about ability varies greatly across different types of managers.

[11] To see that Level 1 is more informative than Level 3, note that $\alpha_1\beta_3 = 0.252 > \alpha_3\beta_1 = 0.248$ and $(1-\alpha_1)(1-\beta_3) = 0.2479 < (1-\alpha_3)(1-\beta_1) = 0.2483$. To see that Level 3 is more informative than Level 2, note that $\alpha_3\beta_2 = 0.267 > \alpha_2\beta_3 = 0.266$ and $(1-\alpha_3)(1-\beta_2) = 0.232 < (1-\alpha_2)(1-\beta_3) = 0.233$.

For example, for managers of the first skill type, it takes 24 years of consecutive high signals at Level 1 for the initial prior to converge to 0.90, 29 years at Level 2, and 28 years at Level 3. For managers of the fourth skill type, that process takes only 15 years at Level 1, 18 years at Level 2, and 17 years at Level 3. Thus, the estimated heterogeneity in initial priors, together with the different informativeness of jobs, has a major impact on the speed of learning for different manager skill types assigned to different levels.

These results about the speed of learning are consistent with the findings of Nagypál (2007). Nagypál estimates a learning model of labor market turnover that focuses on firm-specific ability and abstracts from individuals' ex ante heterogeneity and within-firm mobility. She also finds that learning occurs slowly over time. Specifically, from Figure 7 in her paper (where all plots are truncated at the tenth year of tenure), the convergence of beliefs, as reflected in the tenure profile of match quality and output, seems to occur past the tenth tenure year. In this sense, our results are similar: we both find that learning occurs gradually over time.

Finally, I document an interesting pattern for the error in recorded ratings. From the estimated values of d_0, $d_2(1)$, and $d_2(2)$ in Table 10, it is possible to show that the probability of recording too high a rating (that is, recording low performance as high) significantly decreases with tenure, whereas the probability of recording too low a rating (that is, recording high performance as low) substantially increases with tenure.

A potential explanation for the first finding is that during the initial years of employment at the firm, unsuccessful performance may be better tolerated as ascribed to lack of experience. But as tenure accumulates, managers' career prospects at the firm may be severely compromised due to poor performance. The seriousness of performance evaluation for managers with unsatisfactory performance thereby increases with tenure, leading to fewer mistakes in recording low performance as tenure accumulates. A potential explanation for the second finding is that as uncertainty decreases over time, ratings become less critical for managers who are known to have performed satisfactorily and, thus, their performance is evaluated more haphazardly, leading to more mistakes in recording high performance as tenure accumulates. Lastly, note that the significance of the estimates of the coefficients on covariates in the error rates implies that recorded error leads to bias.

Productivity and Technological Human Capital. I turn now to discussing the estimates of the parameters governing productivity and human capital accumulation that prove most important in explaining the patterns of transition of managers across levels. Recall that these parameters are a_{kt}, $b_{kk_{t-1}t}$, and c_{kt} from (22). The estimates of these parameters have three important implications. First, the estimates imply a substantial cost to demotion. Second, the estimated difference in one-period expected output between a high- and a low-ability manager at each job is hump-shaped with tenure. As such, it is largely responsible for the observed hump-shaped pattern of the hazard rate of promotion to Levels 2 and 3. Third, the transferability of acquired human capital differs substantially across levels: the human capital acquired at Level 3 is highly specific to Level 3, whereas the human capital acquired with experience at the firm is highly valuable at both Levels 2 and 3.

First, consider how the parameters imply a substantial cost of demotion as captured by the parameters $b_{kk_{t-1}t}$. Recall that $b_{kk_{t-1}t}$ measures the change in one-period expected output at Level k in tenure t for a manager who was assigned to Level k_{t-1} in tenure $t-1$. Note that the parameters b_{124} and b_{125} are negative. This implies that if a manager is assigned to Level 1 in the fourth or fifth year of tenure and was assigned to Level 2 in the previous period, then the firm incurs a loss in output by demoting the manager. Similarly, the parameters b_{126} and b_{127}, also negative, help account for the lack of demotions from Level 2 to Level 1 at higher tenures. (Since the parameters b_{125}, b_{126}, and b_{127} have proved to be insignificantly different from each other, I have set them equal and in Table 10 just report b_{125}. See the Appendix for details.)

One interpretation of these parameters is that they capture reallocation costs for the firm, due, for instance, to the reorganization of production and, potentially, to the retraining of a manager for Level 1 tasks after the assignment to Level 2. A related interpretation is that these parameters measure the loss of human capital that is specific to Level 2. To see this, note that the productivity loss from demotion from Level 2 to Level 1 in the fourth and fifth year of tenure, measured by $b_{124} - b_{224}$ and $b_{125} - b_{225}$, are effectively equal to b_{124} and b_{125}, since b_{224} and b_{225} have proved insignificantly different from zero. The same argument applies to the sixth and seventh year of tenure.

Consider now the magnitudes of these losses associated with demotions. Since, as explained in the Appendix, I normalize the expected output of a manager in the first year of tenure at 1,000, the parameters b_{124} and b_{125} imply a sizeable cost of demotion, decreasing in tenure at the firm, of 70.5 percent (from 704.735/10) in the fourth year of tenure and 48.0 percent (from 479.607/10) between the fifth and seventh years of tenure, relative to the expected output of a manager in the first year of tenure.

Second, the estimates imply that the difference in one-period expected output between a high- and a low-ability manager at each job is hump-shaped with tenure. The relevant parameters are c_{kt}, which, by (22), measure the difference in one-period expected output at Level k and tenure t between a manager known to be of high ability ($p_{it} = 1$) and a manager known to be of low ability ($p_{it} = 0$). At each level, these parameters at first increase and then decrease with tenure. At Level 1, the peak occurs in the second year of tenure; at Level 2, in the second and fourth years; and at Level 3, in the seventh year.

Crucially, this hump-shaped pattern of the differences in expected output between managers of high and low ability helps the model account for the observed hump-shaped pattern of the hazard rates of promotion at Levels 1 and 2. The bulk of promotions occurs in the data in intermediate tenure years, when, according to the model, the gap in acquired skills between managers of high and low ability is largest. By promoting those managers perceived to be most able, the firm selects not only managers most likely to be of high ability but also those managers who have acquired the greatest amount of new skills. Hence, the firm's incentive to promote is the highest.

Third, these estimates imply that the transferability of acquired human capital differs substantially across levels. Consider the parameters b_{33t}, which measure the change in one-period expected output at Level 3 in tenure t for a manager assigned to Level 3 in the previous period. From the fourth year of tenure on, the parameters b_{33t} are all positive, large, and very precisely estimated. Instead, the parameters b_{k3t} are zero for $k \neq 3$. These two findings imply a significant degree of specificity of the human capital acquired at Level 3. This result is consistent with the analysis of job titles and descriptions by BGH, who argue that tasks performed at Level 3 are markedly different from those performed at Levels 1 or 2. (See the Supplementary Appendix.) Note that this specificity of acquired human capital at Level 3 helps the model account for the high retention rate of managers at this level at high tenures and for the lack of demotions from Level 3 to Level 2. In addition, the fact that wages are on average higher at Level 3, as discussed later, implies that the human capital acquired at Level 3 is also transferable to other firms. Hence, my estimates imply that human capital acquired at Level 3 at medium and high tenures is task-specific but general across firms.

Finally, I document that the human capital acquired with experience at the firm is highly valuable at Levels 2 and 3. To see this, consider again the parameters c_{kt}. In my parameterization of c_{kt}, I allow for a common component to these parameters across Levels 2 and 3 (γ_{22}) as well as tenure- and level-specific components. This common component is estimated to be large. For example, the common component accounts for 72.5 percent (from γ_{22}/c_{22} and γ_{22}/c_{24}) of the output gap between managers of high and low ability at Level 2 in the second and fourth years of tenure, 82.3 percent (from γ_{22}/c_{34}) of this gap at Level 3 in the fourth year of tenure, and 100 percent (from γ_{22}/c_{35} and γ_{22}/c_{36}) of this gap at Level 3 in the fifth and sixth years of tenure. (See the Appendix for details.)

Turnover. The estimation results imply that the exogenous separation shocks η_{kt} are a primary determinant of managerial turnover. These rates are very precisely estimated at each level and tenure, are slightly decreasing in tenure at Levels 1 and 2, constant at Level 3, and closely track the tenure profile of the empirical hazard rate of separation at each level. Altogether, these facts imply that the observed proportion of managers separating from the firm, increasing in tenure, is due to a steady accumulation of random outside opportunities by managers or demand shocks by the firm that affect its ability to retain managers.

These findings are also consistent with the evidence of Baker, Gibbs, and Holmström (1994b,

p. 931), who note that empirical exit rates are insensitive to a manager's relative wage within a level. Based on this fact, they conclude that the selection effect of retention may not be that important. This conjecture is supported by my estimates. Baker, Gibbs, and Holmström (1994a,b) also suggest that, as a consequence, human capital may be transferable across firms. As I discuss next, my estimates of the wage parameters indeed support the idea that human capital is transferable across firms.

Wages. Here I discuss the implications of the parameter estimates for the importance of observable managerial characteristics, job levels, year of entry, tenure, and unobservable characteristics and outcomes (priors, productivity shocks, and measurement error) for paid wages. Overall, these estimates are consistent with several findings in the literature about compensation in firms and wage growth with tenure despite my data being limited to one firm.

(a) Observable Characteristics. Estimates of the coefficients on entry age and education imply that these characteristics quantitatively matter for differences in wages across managers. Recall that the estimates of these coefficients capture the importance of the human capital acquired before entry into the firm for the firm's paid wages. The fact that these coefficients, as expected, are positive and significant imply that human capital acquired before entry into the firm is transferable to other firms. Specifically, at Levels 1 and 2, the effect of an additional year of age, evaluated at the average age at entry of 29.71 years, is 1.0 percent (from $\varpi_1 + 2\varpi_2(29.71) = 0.028 - 2(0.0003)29.71 = 0.0102$), whereas the effect of an additional year of schooling is 2.2 percent (from $\varpi_3 = 0.022$). At Level 3 the corresponding numbers are 0.4 percent (from $\varpi_{13} + 2\varpi_{23} = 0.010 - 2(0.0001)29.71 = 0.0041$) and 2.1 percent (from $\varpi_{33} = 0.021$).

These estimates of the effects of age and education on wages are comparable to analogous findings in the literature. For instance, Belzil and Bognanno (2008, Table 1) estimate coefficients of 0.0127 and 0.0494 for the (just linear) impact of age and education on (log) wages in a large multi-firm sample of U.S. executives observed between 1981 and 1988, a window of observation very similar to mine. Not surprisingly, since I restrict attention to a highly educated group of individuals, the effect of education on wages that I estimate is smaller than that of Belzil and Bognanno (2008).

(b) Wages at Different Levels. According to the model, promotions lead to sizeable increases in wages. To see this, note that a promotion from Level 1 to Level 2 implies an increase in annual wages of \$781, whereas a promotion from Level 2 to Level 3 implies an increase in annual wages of \$5,723. These figures are obtained by first averaging across manager skill types the level-specific intercepts, given by $\sum_i q_i\varpi_{0i1} = 9.054$, $\sum_i q_i\varpi_{0i2} = 9.141$, and $\sum_i q_i\varpi_{0i3} = 9.619$, and then computing the differences in these values after converting them back from logs to levels.

These results are in line with much of the literature on internal labor markets. For example, a robust finding in this literature, as reviewed by Gibbons and Waldman (1999a,b) and Waldman (forthcoming), is that promotions entail substantial increases in pay. More specifically, my results are comparable to those obtained by Belzil and Bognanno (2008), who find that promotions by one level are associated with an increase of \$1,352 or \$1,598 in wages, according to their respective fixed effects and ordinary least squares estimates, after controlling for changes in firm profits, sales, and size. Indeed, their estimated promotion premium is between my two estimates for the wage premium associated with a promotion from Level 1 to Level 2 and from Level 2 to Level 3. Finally, my estimates imply that wages are quite convex in job levels. Indeed, the average wage increase resulting from a promotion from Level 2 to Level 3 is more than five times larger than the average wage increase resulting from a promotion from Level 1 to Level 2. Such convexity is in line with similar findings on promotion premia increasing in the job level documented in other studies, including that of Belzil and Bognanno (2008) and, naturally, BGH.

(c) Year of Entry Effects. I find quantitatively important effects of a manager's year of entry into the firm on average wages. In the model, these effects are captured by the year-of-entry dummies for entry between 1974 and 1975 (the coefficient ϖ_{y5}) and 1979 (the coefficient ϖ_{y9}). (The baseline years are 1970, 1971, 1972, and 1973. I also restricted the coefficient on the year-of-entry dummy for 1975 to be equal to the one for 1974, given that the difference between the two is insignificant.) The coefficients are all (precisely) estimated to be negative. This finding is consistent with evidence on the tightness of the labor market in the recession years between 1974 and 1979, which in my data depresses the average wages of managers who entered into the firm during those years relative to those who entered in earlier years. More generally, this result highlights the importance for wages of external labor market conditions at the time of entry into a firm, in particular business cycle effects. (See also the discussion of Baker, Gibbs, and Holmström (1994b) on this point.)

(d) Tenure, Priors, Productivity Shocks, and Measurement Error. Consider now the rest of the terms in the wage equation (30). I document three main findings concerning the effect of tenure and unobservables on paid wages.

First, the tenure effect is estimated to be non-zero only at Level 1, and it essentially decreases with tenure, thus helping the model account for the decrease with tenure of the average wage of managers continuously retained at Level 1. Hence, overall the dynamics of wages is driven by learning, human capital acquisition, and competition rather than an exogenous tenure or experience trend.

Second, consider the coefficients on priors. Recall that the coefficient on the first-degree term in the prior is ω_{2i}. (The coefficient ω_3 on the second-degree term in the prior, which arises from the nonlinear component of the learning and human capital premium in wages, is estimated to be not significantly different from zero.) Note that the coefficients ω_{2i} capture the sum of two components: the effect of the prior on the expected output of a manager at the best competitor of my firm and the effect of the prior on the linear part of the learning and human capital premium in wages. These coefficients do vary substantially across manager skill types. Namely, the coefficients $\{\omega_{2i}\}_{i=1}^{4}$ are estimated to be $(2.371, 1.833, 1.316, 1.364)$. These magnitudes imply that for a given increase in perceived ability, captured by a marginal change in the prior, the resulting increase in (the mean log of) wages for the first type of managers is almost twice as large as the increase for the fourth type.

Third, consider the last term u_{inkt}, which is the sum of the productivity shock, ε_{inkt}, and measurement error, ε^{m}_{inkt}. Note that this residual unobserved source of variability in wages essentially decreases with the level. Specifically, for all four types, the residual variance σ_{ik} is higher at Level 1 than at Level 2, and for the first three manager types, it is higher at Level 1 than at Levels 2 and 3, whereas for the first two manager types, it is higher at Level 2 than at Level 3. Despite this pattern, the model implies that the variance of wages is higher at higher levels as in the data (the estimated standard deviation of wages at Level 1, Level 2, and Level 3, pooled across tenures, is \$6,936 at Level 1, \$7,077 at Level 2, and \$8,046 at Level 3). Thus, the model's prediction that the variance of wages is higher at a higher level than at a lower level is generated by the endogenous mechanisms of the model, such as the increased dispersion of informational and technological human capital with tenure, rather than by unexplained disturbances.

Implications for Market Power. Based on the estimates of the parameters for expected output and wages at each level, I can conclude that my firm has market power: the correlation coefficient between average expected output and average wages over the first seven years of tenure is 0.478 at Level 1, 0.380 at Level 2, and 0.521 at Level 3 (excluding the third tenure year, in the case of Level 3; the value is otherwise much lower). Note that if my firm behaved as a perfect competitor, then such correlation would be close to one. This finding confirms the importance of explicitly modeling imperfect competition in the labor market to empirically assess the determinants of paid wages and wage growth on the job.

Implications for Within-Job Wage Growth. Overall, the estimates imply an increase in wages of 19.4 percent over the first seven years of tenure at the firm (for managers continuously employed at the firm), corresponding to an average yearly growth rate of 3.2 percent. This estimated magnitude for wage growth on the job is consistent with the estimates of within-job wage growth by Topel (1991) (see his Table 2) and is bracketed by those of Buchinsky, Fougère, Kramarz, and Tchernis (2010) (see their Figure 2). Specifically, based on their estimates from the Panel Study of Income Dynamics (PSID), Buchinsky et al. (2010) document that yearly wage growth is between 2.9 percent and 8.7 percent for individuals with a college degree. Thus, my estimate of 3.2 percent lies between their two estimates.[12]

Importantly, the increase in wages I measure is not simply generated by an exogenous tenure trend, but rather arises endogenously from competition between firms in the labor market. (As discussed, a manager's age and education do not contribute to the dynamics of wages.) Namely, competition between firms leads wages to reflect the portion of a manager's informational and technological human capital accumulated at my firm that is also valuable to other firms. In this sense, my exercise can be interpreted as providing a micro-foundation for the process of wage growth on the job that accounts for the endogeneity of job mobility within and between firms. This micro-foundation is in line with one of the main insights from Topel (1991) and Altonji and Williams (2005), among others, that individual heterogeneity is an important factor of the dynamics of jobs and wages.

This approach is also consistent with the one of Buchinsky et al. (2010). These authors consider a decision-theoretic framework in which workers search for jobs offering the highest (expected present discounted) value of consumption and derive the optimal decision rules for employment with no mobility between firms, employment with mobility between firms, and nonparticipation. My results confirm their insight that accurate estimates of the wage process require explicitly controlling for job-specific components in the wage equation. Buchinsky et al. (2010) do so by introducing in the wage equation a function of a worker's experience and tenure in past jobs that is meant to capture the overall effect of a worker's career in the labor market on the worker's wage. In contrast to their work, I view my exercise as a first attempt to provide a theory for the wage process that incorporates the effect of mobility within firms based on the dynamics of informational and technological human capital acquisition.

Lastly, my exercise has implications for the importance of the accurate measurement of career outcomes within firms for quantifying wage growth on the job. A consensus has emerged in the labor economics literature that imperfect measurement of the timing of job and wage changes across firms may account for the large differences in the estimates of the returns to tenure based on survey data. (See Altonji and Williams (2005), Dustmann and Meghir (2005), and Buchinsky et al. (2010).) In my administrative dataset, by construction, the timing of job and wage changes within a firm are well measured. Based on these data, I find that important moments of the distribution of wages vary with the job level and tenure at a level. This variation and the implied nonlinearity in the yearly growth rate of wages suggest that the estimation

[12] To deduce these yearly wage growth percentages, I converted Buchinsky et al. (2010)'s cumulative log point wage growth over an 18-year period into a yearly growth rate of levels. Specifically, they compare the wage growth for two types of hypothetical workers with distinct career paths. One group consists of individuals working through the entire sample period in one job, whereas the other includes those working at one firm for the first 4 years and then changing to a new job in which they stay for the remainder of the sample period. In both groups, they focus on new entrants, that is, individuals with 5 years of experience and 2 years of tenure, and experienced workers, that is, individuals with 15 years of experience and 6 years of tenure. They document that, over the 18-year period they analyze, wage growth for individuals with a college degree is 0.94 log points in the case of new entrants and 0.42 log points in the case of experienced workers. These findings imply a yearly growth rate between 2.9 percent (from $[\exp(0.42) - 1]100/18$) and 8.7 percent (from $[\exp(0.94) - 1]100/18$).

of returns to tenure requires accurate dating not only of the timing of job and wage changes *between* firms, as the literature has suggested, but also of the timing of job and wage changes *within* firms.

Overall, the similarities between my estimates of various aspects of the wage process, such as the return to entry age, the size of wage increases at promotion, the convexity of wages in job levels, and wage growth on the job, with estimates of these aspects in the literature suggest that my data display features consistent with many well-documented characteristics of managerial careers.

6 The Roles of Learning, Experimentation, and Persistent Uncertainty

In this section I use the estimated model to assess the roles of learning, experimentation, and persistent uncertainty about ability in accounting for the observed patterns of job and wage mobility in my data. I do so through three sets of experiments. To assess the role of learning, I compare the outcomes in the model, which I now refer to as the *baseline model*, to those arising without learning, that is, when jobs at all levels are uninformative about ability. To assess the role of experimentation, I compare the outcomes in the baseline model to those arising when all jobs are equally informative about ability, so that no experimentation is possible. Lastly, to assess the role of persistent uncertainty about ability, I compare outcomes in the baseline model to those arising when uncertainty is resolved more quickly through faster learning. Overall, I find that learning, experimentation, and persistent uncertainty are key factors in generating observed job and wage mobility.

The Role of Learning. To evaluate the role that learning plays in generating the patterns of job and wage mobility in the baseline model, I consider what these patterns would look like if learning were absent. Specifically, I consider a *no learning* version of the model in which all jobs are assumed to be uninformative about ability (that is, β_k at each Level k equals the estimated value of α_k, whereas the remaining parameters are left unchanged). I find that several patterns are quite different when learning is absent from the model. In particular, the absence of learning leads to a smaller wage growth with tenure and slower promotions.

In Table 11A, I compare some statistics for wages in the two models. The second column of the table shows average wages by level pooled across the first seven years of tenure (the first three rows), the standard deviation of wages by level pooled across these tenures (the next three rows), and the cumulative wage growth by tenure pooled across levels (the next six rows) in the baseline model. The last row reports cumulative wage growth over the first seven years at the firm only for managers continually employed at the firm. Note that, even in the baseline case, this growth rate is higher than that computed when separating managers are included: this difference in wage growth suggests the presence of selection through retention. The third column shows the same statistics in the no learning version of the model.

One role of learning is to generate higher wage growth as tenure accumulates. For instance, over the course of the first seven years at the firm, learning contributes to 26 percent of measured wage growth (from $(19.4-15.4)100/15.4 = 26.0$ percent) relative to the model without learning. Learning also helps account for the observed variability of wages at each level. For instance, learning adds 23 percent to the variability of wages at Level 3 (from $(8,046-6,534)100/6,534 = 23.1$ percent).

Table 11B displays the distribution of managers across levels by tenure in the two models. Clearly, learning also helps account for the mobility of managers across levels. For example, in the baseline model, by the third year of tenure about 9 percent of managers have been promoted to Level 3, whereas in the no learning model fewer than 1 percent have. By the seventh year of tenure, in the baseline model over 32 percent of managers have been promoted to Level 3, whereas in the no learning model only about 24 percent have.

Thus, overall learning accounts for a sizeable fraction of observed wage growth, wage variability, and job mobility at the firm.

The Role of Experimentation. Experimentation through job assignment also has a substantial impact on managers' careers at the firm. To measure the role that experimentation plays in generating the patterns of job and wage mobility in the baseline model, I consider what these patterns would look like if all jobs were equally informative, so that no experimentation can take place. Somewhat surprisingly, this lack of experimentation has almost no effect on wages. Instead, preventing experimentation has a substantial effect on job mobility.

Here I focus on one experiment, referred to as *equal informativeness as Level 1.* In this experiment, I suppose that jobs at Levels 2 and 3 are as informative as at Level 1 by setting α_k and β_k for Levels 2 and 3 equal to their estimated values for Level 1 while leaving the rest of the parameters unchanged. I report results for the analogous experiments for Levels 2 and 3 in the Supplementary Appendix (Tables A.12–A.14). Since the patterns of wages for this experiment are remarkably similar to the patterns in the baseline model, I also relegate those results to the Supplementary Appendix (Table A.12) and focus here on the implications of the experiment for job mobility.

Table 12 shows that without experimentation, nearly all of the managers who do not separate from the firm are quickly assigned to Level 3 and are retained there. Indeed, apart from the first year of tenure, the proportion of managers assigned to Level 1 is very small in each year.

Remarkably, this pattern holds even though learning occurs very slowly over time, as discussed. This experiment thus highlights an important point. In the experiment, managers are rapidly promoted to high levels even though the process of information acquisition takes place gradually with tenure at the firm. Hence, according to the model, the speed of promotion is not informative about the absolute speed of learning in jobs but rather about the relative informativeness of performance at different jobs, and, thus, about the scope for experimentation. This effect proves to be quantitatively very important for the dynamics of job mobility with tenure at the firm.

The Role of Persistent Uncertainty. As discussed, in the baseline model, learning occurs very gradually over time. Hence, uncertainty about managers' ability persists even after several years of tenure of a manager at the firm. To assess the role that this persistent uncertainty plays in generating the patterns of job and wage mobility in the baseline model, I consider what these patterns would look like if this uncertainty were resolved more quickly. It turns out that faster learning leads to higher wage growth, greater dispersion in wages, and faster promotions.

Specifically, I conduct two experiments. In the *fast learning at Level 1* case, I suppose that jobs at Level 1 are nearly perfectly informative about ability by setting $\alpha_1 = 0.99$ and $\beta_1 = 0.01$ while leaving the rest of the parameters unchanged. In the *fast learning at Level 2* case, I suppose that jobs at Level 2 are nearly perfectly informative about ability by setting $\alpha_2 = 0.99$ and $\beta_2 = 0.01$ while leaving the rest of the parameters unchanged.

Consider the case of fast learning at Level 1. Table 11A shows the implications of this experiment for wages, and Table 11B shows the implications for job assignment. To see that fast learning leads to higher wage growth with tenure, note that cumulative wage growth by the second year of tenure for this case is much higher than in the baseline model, 39.3 percent here compared to only 4.6 percent in the baseline model. Over the first seven years of tenure at the firm, this faster learning leads wages to grow more than 60 percent compared to approximately 20 percent in the baseline model. To see that fast learning increases the dispersion in wages, note that the standard deviation of wages at the three levels in this case is substantially larger than in the baseline model. For instance, the dispersion in wages paid to managers at Level 3 is over five times larger than in the baseline model. Finally, to see that faster learning leads to faster promotions, note that in this case managers are promoted to Level 3 more quickly than in the baseline model. For instance, as shown in Table 11B, by the third year of tenure, 20 percent of managers are assigned to Level 3, whereas in the baseline model only about 9 percent are.

Now consider the case of fast learning at Level 2. To see that fast learning at Level 2 also leads to higher wage growth, note in Tables 11A and 11B that wage growth is dramatically higher from the third year of tenure on than in the baseline model. Clearly, this fast learning also leads to much larger wage dispersion at Level 2 but, perhaps surprisingly, much lower wage dispersion at Level 3 relative to the baseline model. I will comment later on this finding about the impact of fast learning at Level 2 on wage dispersion at Level 3. Finally, this faster learning implies faster promotions. For example, as shown in Table 11B, this fast learning also leads to more than twice the number of managers assigned to Level 2 by the second year of tenure and to a much higher proportion of managers assigned to Level 3 in the third and fourth years of tenure than in the baseline model. Faster learning at Level 2, however, leads to a lower percentage of managers assigned to Level 3 at higher tenures than in the baseline model.

Overall, persistent uncertainty about ability tends to substantially compress the levels and growth rate of wages over time: cumulative wage growth is between 32 and 43 percentage points higher when, respectively, the second-lowest and lowest hierarchical levels become almost perfectly informative about ability.

To understand why at high tenures many managers are still assigned to Level 2, note that for managers known to be of high ability, Level 2 is a productive assignment under this scenario of fast learning at Level 2. The reason is that at any given prior, increasing α_2 increases a manager's expected output at Level 2 whereas decreasing β_2 decreases expected output at Level 2. As is apparent from (2), at high enough priors, the first effect dominates the second. Thus, compared to the baseline model, increasing α_2 and decreasing β_2 makes Level 2 more attractive for managers with high priors, who tend to be managers with high tenures at the firm. This effect is pronounced enough that Level 2 is a more productive assignment than Level 3 for managers with high priors at high tenures. In turn, that most managers at Level 3 have low priors also implies that the standard deviation of wages at Level 3 is much lower than at Level 2—indeed, lower than in the benchmark model.

7 Conclusion

In this paper, I have developed and estimated an equilibrium model of the labor market that integrates learning, job assignment, and human capital acquisition to provide a behavioral foundation for the stochastic structure of job and wage changes characteristic of careers in firms. My results show that learning and experimentation are prime contributors to the patterns of observed job and wage mobility and that persistent uncertainty about ability is a quantitatively important determinant of the dynamics of wages with tenure. Naturally, as for any structural model, the insights the model has offered must be subject to further verification. Yet, my overall findings attest to the potential of learning models, augmented along the dimensions considered, to capture salient features of careers in firms.

In the paper, I have showed that primitive parameters of interest can be recovered solely on the basis of a sufficiently rich best-response problem evaluated at equilibrium, which naturally provides the model's main estimating equations. A similar approach may be useful in other contexts of limited data availability or in which equilibrium distributions of outcomes of interest cannot conveniently be empirically exploited.

Although the matching process between workers and firms has been extensively studied in the labor economics literature, much less is known about the matching between workers and jobs within a firm. My exercise contributes to an empirical literature still in its infancy, the goal of which is to shed light on the internal working of firms and bridge several areas of study: labor economics themes on the returns to firm tenure and labor market experience, industrial organization themes on the activities and boundaries of firms, personnel economics themes on the importance of firm strategies for hiring and retaining talented individuals (see the review by Oyer and Shaefer (2011)), and, lastly, human resource management themes on the relationship between human resource practices and productivity (see the review by Bloom and Van Reenen (2011)).

In this paper, due to data limitations, I have focused on job and wage mobility within a firm. The model proposed, however, can produce rich patterns of job and wage mobility between firms as well. A promising avenue of future research would, therefore, be to explore the extent to which a version of the model that incorporates search and matching frictions can explain broad patterns of evidence about individual and aggregate outcomes of careers in firms and turnover between firms.

A Appendix: Details on the Data and the Model

Here I provide further details about the estimation sample and the empirical specification of the model in terms of the parameters governing level assignments, exogenous separations, and paid wages. I conclude by presenting evidence on model specification based on Pearson's χ^2 test.

A.1 Estimation Sample

In estimation I restrict attention to individuals entering the firm between 1970 and 1979 for two reasons. First, to avoid potential censoring problems for individuals first observed in 1969, I consider individuals entering from 1970 on. In the data, in fact, it is not possible to distinguish whether new entrants into managerial positions in any given year are also new hires at the firm. For instance, a manager could be promoted from a clerical to a managerial position and still be recorded as an entrant. Second, to allow for variability in managers' year of entry and a sufficiently long window of observation for each manager, I exclude entrants after 1979. The specific choice of 1979 is motivated by reasons of comparability of my results with those of BGH, who primarily focus on these years for their longitudinal analysis.

Note that the BGH data contain information on managers' salary (or base pay, which I referred to in the paper simply as wage) and bonus pay, all expressed in 1988 constant U.S. dollars. However, data on bonuses before 1981 are not available. Overall bonuses are paid to 25 percent of employees in these later years, mainly to managers at the highest levels. Also, for most managers, bonuses do not significantly affect total compensation: the median bonus for those managers receiving one at (the original) Levels 1–3 in the data is less than 10 percent of salary; for those at (the original) Level 4, it is less than 15 percent. (See Gibbs (1995) for an analysis of these data.) Overall, whereas salary information over the first ten years of tenure at the firm is missing for fewer than 13 percent of employed managers in each sample year, bonus information is missing for no fewer than 45.8 percent of managers with much higher percentages at low tenures. Respectively, the percentages missing are 100, 100, 85.1, 70.8, 63.0, 56.0, and 45.8. (Including observations with a recorded bonus of zero, these percentages between the third and seventh years of tenure increase to 96.0, 89.5, 85.9, 81.0, and 76.7, respectively.)

An additional reason for the exclusion of bonus data from the estimation sample is that the data display evidence of a *bonus list*, in the sense that almost all managers who receive a bonus in one of the first six tenure years also receive a bonus in each subsequent year, seemingly regardless of (recorded) performance. Accordingly, the assumptions I maintain in estimation are that total compensation is separable in base and bonus pay and that the expected bonus payment, at the time a manager accepts an employment offer, is zero. See the Supplementary Appendix for further details on the construction of the estimation sample.

A.2 Model Parameterization and Fit

I first discuss omitted details of the parameterization of the model and then present results on model fit. Altogether, the model has 75 estimated parameters, 38 parameters pertaining to managers' transitions across levels, including the parameters governing the initial distribution of prior beliefs and performance ratings, and 37 pertaining to wages.

Productivity and Technological Human Capital Parameters. Recall that, as discussed, the parameters a_{kt}, $b_{kk_{t-1}t}$, and c_{kt} are to be interpreted as differences between the values of the corresponding parameters at my firm and at the second-best firm. So, whenever any of the parameters a_{kt}, $b_{kk_{t-1}t}$, and c_{kt} for expected output at job k in period t at my firm are found to be not significantly different from zero, this implies that any such parameter of expected output turned out to be the same across firm A and the second-best firm.

Notice that, without other restrictions, the specification of expected output at my firm from (22) leads to more than 100 parameters to be estimated. Given the lack of direct information on output in my data, I conserve on parameters in two ways. First, I set to zero parameters that turn out to be quantitatively insignificant, when constraining them to be equal to zero does not affect any other parameter estimate. As discussed, any such case is to be interpreted as an instance in which a certain productivity parameter has the same value at firm A and at the second-best firm. Second, when differences in certain parameters across tenures for the same level or across levels, either for the same tenure or different tenures, are quantitatively insignificant, and have no effect on any other parameter estimate, I set the parameters equal to each other. I also specified each parameter c_{kt} as containing a level-specific component, a tenure-specific component, a component general to other levels, and a component common to $b_{kk_{t-1}t}$ across some levels and tenures.

Recall that I have suppressed the firm subscript. Based on these observations, I obtain the following parameterization. A manager's expected output at Level 1 is given by

$$y(p_{it}, t-1, k_{t-1}, 1) = a_{11}I(t=1) + c_{12}p_{i2}I(t=2) + b_{124}I(k_3=2)I(t=4) + b_{125}I(k_4=2)I(t=5)$$
$$+b_{126}I(k_5=2)I(t=6) + b_{127}I(k_6=2)I(t=7)$$

with $a_{11} = 1,000$ (recall that all managers are assigned at entry to Level 1 so any large value for a_{11} would produce the same initial probability of assignment to Level 1), and $b_{127} = b_{126} = b_{125}$. At Level 2, expected output is given by

$$y(p_{it}, t-1, k_{t-1}, 2) = c_{22}p_{i2}I(t=2) + c_{23}p_{i3}I(t=3) + c_{24}p_{i4}I(t=4) + c_{25}p_{i5}I(t=5)$$
$$+c_{26}p_{i6}I(t=6) + c_{27}p_{i7}I(t=7) + c_{28}p_{i8}I(t=8)$$

with $c_{22} = \gamma_{22} - b_{124}$, $c_{23} = \gamma_{23} - b_{124}$, $c_{24} = c_{22}$, $c_{25} = \gamma_{25} - b_{125}$, $c_{26} = \gamma_{26} - b_{125}$, $c_{27} = c_{26}$, and $c_{28} = \gamma_{28} - b_{125}$ (γ_{28} turned out not significantly different from zero and b_{125} negative). Note that the specification of c_{2t} between $t = 3$ and $t = 8$, which has proved relevant at the estimated parameter values, does not impose any restriction between the value of these parameters at Levels 2 and 3. In estimation, γ_{22} has proved to be the only common component across Levels 2 and 3. To see this, note that at Level 3 expected output is given by

$$y(p_{it}, t-1, k_{t-1}, 3) = c_{31}p_{i1}I(t=1) + c_{32}p_{i2}I(t=2) + c_{33}p_{i3}I(t=3)$$
$$+[b_{334}I(k_3=3) + c_{34}p_{i4}]I(t=4)$$
$$+[b_{335}I(k_4=3) + c_{35}p_{i5}]I(t=5) + [b_{336}I(k_5=3) + c_{36}p_{i6}]I(t=6)$$
$$+[b_{337}I(k_6=3) + c_{37}p_{i7}]I(t=7) + c_{38}p_{i8}I(t=8)$$

with $c_{33} = c_{32} = c_{31}$, $c_{34} = \gamma_{22} - c_{31}$, $c_{36} = c_{35} = \gamma_{22}$, and $b_{336} = b_{335}$. So the estimated output parameters are b_{124}, b_{125}, c_{12}, γ_{22}, γ_{23}, γ_{25}, γ_{26}, b_{334}, b_{335}, b_{337}, c_{31}, c_{37}, and c_{38}.

Exogenous Separation Parameters. Next consider the separation shocks that managers at my firm face. Note that in the model these shocks are allowed to depend flexibly on tenure at the firm and current level assignment. In the spirit of parsimony, I allow only for variation in these shocks across levels and tenures that proves statistically significant, whenever setting these parameters equal across levels or tenures does not affect any other parameter estimate. As a result, the parameters of the probability of a separation shock, respectively, at Levels 1, 2, and 3 are: (1) at the end of period 1: η_{11}, η_{21}, and η_{31}; (2) at the end of period 2: none, since $\eta_{12} = \eta_{11}$, $\eta_{22} = \eta_{21}$, and $\eta_{32} = \eta_{31}$; (3) at the end of period 3: ξ_3, since $\eta_{13} = \eta_{14} + \xi_3$ (see the parameterization for the next tenure to understand this), $\eta_{23} = \eta_{22} + \xi_3$, and $\eta_{33} = \eta_{32}$; (4) at the end of period 4: η_{14} and η_{24}, since $\eta_{34} = \eta_{24}$; (5) at the end of period 5: η_{25}, since $\eta_{15} = \eta_{14}$ and $\eta_{35} = \eta_{25}$; (6) at the end of period 6: η_{26}, since $\eta_{16} = \eta_{15}$ and $\eta_{36} = \eta_{26}$; (7) at the end of period 7: η_{27}, since $\eta_{17} = \eta_{16}$ and $\eta_{37} = \eta_{27}$; and (8) at the end of period 8: none, since $\eta_{18} = \eta_{17}$, $\eta_{28} = \eta_{27}$, and $\eta_{38} = 0$. The restriction on η_{38} is imposed in order to identify c_{38} since exit at the end of $t = 8$ affects the proportion of retained managers at $t = 9$, who are not part of the estimation sample. So the separation shock parameters to be estimated are $(\eta_{11}, \xi_3, \eta_{14})$ for Level 1, $(\eta_{21}, \eta_{24}, \eta_{25}, \eta_{26}, \eta_{27})$ for Level 2, and η_{31} for Level 3.

Wage Parameters. From the discussion in the body, I set ϖ_{1k}, ϖ_{2k}, and ϖ_{3k}, respectively, the coefficients on age_n, age_n^2 and edu_n, equal at Level 1 and Level 2. I denote their common value by ϖ_1, ϖ_2, and ϖ_3. In terms of the coefficients on the year-of-entry dummies, I also set $\varpi_{ym} = 0$ for $0 \leq m \leq 3$ and $\varpi_{y4} = \varpi_{y5}$, so the estimated parameters for the year-of-entry dummies are $(\varpi_{y5}, \varpi_{y6}, \varpi_{y7}, \varpi_{y8}, \varpi_{y9})$.

As for the remaining coefficients, at Level 1 I assume that the coefficient on tenure is $\omega_{11t} = \omega_{111}I(t < 5) + \omega_{115}[I(t = 5) + (t = 6)]$, with $\omega_{115} = -\omega_{111}$. The coefficients on tenure at Level 2, ω_{12t}, and Level 3, ω_{13t}, have proved not significantly different from zero. The estimated value of ω_3 has also proved not significantly different from zero. Finally, I assume that the standard error of the normal disturbance u_{inkt} does not vary across skill types at Level 3. So, the estimated variance parameters are $\{\omega_{0i1}, \omega_{0i2}, \omega_{0i3}\}_{i=1}^{4}$, ϖ_1, ϖ_2, ϖ_3, ϖ_{13}, ϖ_{23},ϖ_{33}, $\{\varpi_{ym}\}_{m=5}^{9}$, ω_{111}, $\{\omega_{2i}\}_{i=1}^{4}$, $\{\sigma_{1k}, \sigma_{2k}, \sigma_{3k}, \sigma_{4k}\}_{k=1}^{2}$, and σ_3.

Model Fit. One criterion to formally evaluate model fit is Pearson's χ^2 goodness-of-fit test. I perform this test based on the statistic $q\sum_{r=1}^{R}\{[g(r) - \hat{g(r)}]^2/\hat{g(r)}\}$, where $g(\cdot)$ indicates the empirical density function of a given endogenous variable, $\hat{g(r)}$ denotes the maximum likelihood estimate of the density function of that variable, q indicates the number of observations, and R the number of categories considered (not taking into account the fact that the parameters of the model are estimated).

I compare observed and predicted outcomes in terms of the distribution of managers across levels in each of the first seven years of tenure, the distribution of performance ratings at Levels 1 and 2 in each such year of tenure, and the distribution of wages at each level and in each such year of tenure. The results of the test are as follows. In terms of the distribution of managers across Levels 0 (separation) through 3 at each tenure, the χ^2 goodness-of-fit test does not reject the model at conventional significance levels at any tenure. In terms of the hazard rates of separation, retention at a level, and promotion to the next level at each level and tenure, the test does not reject the model at conventional significance levels, apart from the second, third, fourth, and sixth years of tenure at Level 1 and the second and third years of tenure at Level 2. However, the outcome of the test in this case is arguably very much influenced by the small number of observations at Levels 1 and 2 in those tenure years. In terms of the distribution of performance ratings at Levels 1 and 2 at each tenure, the test does not reject the model at conventional significance levels at any tenure. Finally, in terms of the distribution of wages at Levels 1, 2, and 3 at each tenure, the test does not reject the model at conventional significance levels, apart from the third year of tenure at Level 2 and the fourth, fifth, sixth, and seventh years of tenure at Level 3.

An issue in interpreting these results for wages may be the large attrition in the sample, which implies that only a fraction of observations are on managers at Level 3 at high tenures relative to the size of observations on managers in the first year. In the Supplementary Appendix, I present estimation results based on a larger sample that contains observations on entrants into the firm at Levels 2, 3, and 4 as well. There parameters are more precisely estimated, and the model fit proves better. Indeed, for the two specifications I estimate on this larger sample, the model is almost never rejected.

References

Ackerberg, D.A. (2003): "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination," *International Economic Review* 44(3), 1007–1040.

Altonji, J.G., and N. Williams (2005): "Do Wages Rise with Job Seniority? A Reassessment," *Industrial and Labor Relations Review* 58(3), 370–397.

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.

Baker, G., M.C. Jensen, and K.J. Murphy (1988): "Compensation and Incentives: Practice vs. Theory," *Journal of Finance* 48(3), 593–616.

Belzil, C., and B. Bognanno (2008): "Promotions, Demotions, Halo Effects, and the Earnings Dynamics of American Executives," *Journal of Labor Economics* 26(2), 287–310.

Bergemann, D., and J. Välimäki (1996): "Learning and Strategic Pricing," *Econometrica* 64(5), 1125–1149.

Berkovec, J., and S. Stern (1991): "Job Exit of Older Men," *Econometrica* 59(1), 189–210.

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

Bernhardt, D., and D. Scoones (1993): "Promotion, Turnover, and Preemptive Wage Offers," *American Economic Review* 83(4), 771–791.

Blackwell, D. (1951): "*Comparison of Experiments,*" J. Neyman, Ed., Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press.

Bloom, N., and J. Van Reenen (2011): "Human Resource Management and Productivity," in *Handbook of Labor Economics*, Volume IVb, O. Ashenfelter and D. Card, Eds., Amsterdam: North-Holland, 1697–1767.

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.

Ching, A.T. (2010): "A Dynamic Oligopoly Structural Model for the Prescription Drug Market after Patent Expiration", *International Economic Review* 51(4), 1175–1207.

Ching, A.T., T. Erdem, and M.P. Keane (2011): "Learning Models: An Assessment of Progress, Challenges and New Developments," mimeo.

Crawford, G.S., and M. Shum (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica* 73(4), 1137–1173.

Davis, S. (1997): "Sorting, Learning, and Mobility When Jobs Have Scarcity Value: A Comment," *Carnegie-Rochester Conference Series on Public Policy* 46(1), 327–337.

DeVaro, J., and M. Waldman (2012): "The Signaling Role of Promotions: Further Theory and Empirical Evidence," *Journal of Labor Economics* 30(1), 91–147.

Dustmann, C., and C. Meghir (2005): "Wages, Experience and Seniority," *Review of Economic Studies* 72(1), 77–108.

Easley, D., and N.M. Kiefer (1988): "Controlling a Stochastic Process with Unknown Parameters," *Econometrica* 56(5), 1045–1064.

Eckstein, Z., and K.I. Wolpin (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica* 67(6), 1295–1339.

Erdem, T., and M.P. Keane (1996): "Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets", *Marketing Science* 15(1), 1–20.

Farber, H.S., and R. Gibbons (1996): "Learning and Wage Dynamics," *Quarterly Journal of Economics* 111(4), 1007–1047.

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

Ghosh, S. (2007): "Job mobility and Careers in Firms", *Labour Economics* 14 (3), 603-621.

Gibbons, R., L.F. Katz, T. Lemieux, and D. Parent (2005): "Comparative advantage, Learning, and Sectoral Wage Determination", *Journal of Labor Economics* 23(4), 681–724.

Gibbons, R., and M. Waldman (1999a): "Careers in Organizations: Theory and Evidence," in *Handbook of Labor Economics*, Volume IIIB, O. Ashenfelter and D. Card, Eds., Amsterdam: North-Holland, 2373–2437.

——— (1999b): "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.

Gibbs, M. (1995): "Incentive Compensation in a Corporate Hierarchy," *Journal of Accounting and Economics* 19(2–3), 247–277.

Gibbs, M., and W. Hendricks (2004): "Do Formal Salary Systems Really Matter?" *Industrial and Labor Relations Review* 58(1), 71–93.

Greenwald, B. (1986): "Adverse Selection in the Labour Market," *Review of Economic Studies* 53(3), 325–347.

Harris, M., and B. Holmström (1982): "A Theory of Wage Dynamics," *Review of Economic Studies* 49(3), 315–333.

Hausman, J.A., J. Abrevaya, and F.M. Scott-Morton (1998): "Misclassification of the Dependent Variable in a Discrete-Response Setting," *Journal of Econometrics* 87(2), 239–269.

Heckman, J.J., and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica* 52(2), 271–320.

Holmström, B., and J. Tirole (1989): "The Theory of the Firm," in *Handbook of Industrial Organization*, Volume I, R. Schmalensee and R. Willig, Eds., Amsterdam: North-Holland, 61–133.

Hunnes, A. (forthcoming): "Testing the Role of Comparative Advantage and Learning in Wage and Promotion Dynamics", *International Journal of Manpower.*

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

Keane, M.P., and R.M. Sauer (2009): "Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior," *Econometrica* 77(3), 975–991.

Keane, M.P., and K.I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy* 105(3), 473–522.

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.

Lazear, E. (1992): "The Job as a Concept," in *Performance Measurement, Evaluations, and Incentives*, W. J. Bruns, Jr., Ed., Boston: Harvard University Press, 183–215.

Li, J. (2011): "Job Mobility, Wage Dispersion, and Asymmetric Information," mimeo.

Lluis, S. (2005): "The Role of Comparative Advantage and Learning in Wage Dynamics and Intrafirm Mobility: Evidence from Germany", *Journal of Labor Economics* 23(4), 725-767.

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.

——— (1997): "Estimating Models of Dynamic Optimization with Microeconomic Data," in *Handbook of Applied Econometrics*, Volume II, M. Pesaran and P. Schmidt, Eds., Blackwell, 246–299.

Murphy, K.J. (1992): "Performance Measurement and Appraisal: Motivating Managers to Identify and Reward Performance," in *Performance Measurement, Evaluations, and Incentives*, W. Bruns Eds., Boston: Harvard University Press, 37–62.

Nagypál, É. (2007): "Learning by Doing vs. Learning About Match Quality: Can We Tell Them Apart?" *Review of Economic Studies* 74(2), 537–566.

Oyer, P., and S. Shaefer (2011): "Personnel Economics: Hiring and Incentives," in *Handbook of Labor Economics*, Volume IVb, O. Ashenfelter and D. Card, Eds., Amsterdam: North-Holland, 1769–1823.

Pastorino, E. (2011): "Individual and Market Experimentation with Multi-Armed Dependent Bandits," mimeo.

Prendergast, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature* 37(1), 7–63.

Prescott, E. C., and M. Visscher (1980): "Organization Capital," *Journal of Political Economy* 88(3), 446–461.

Ricart i Costa, J. (1988): "Managerial Task Assignment and Promotions," *Econometrica* 56(2), 449–466.

Rosen, S. (1982): "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics* 13(2), 311–323.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica* 55(5), 999–1033.

——— (1994): "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics*, Volume IV, R. Engle and D. McFadden, Eds., Amsterdam: North-Holland, 3081–3143.

Sauer, R.M. (1998): "Job Mobility and the Market for Lawyers," *Journal of Political Economy* 106(1), 147–171.

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.

——— (1990): "Up-or-Out Contracts: A Signaling Perspective," *Journal of Labor Economics* 8(2), 230–250.

——— (1996): "Asymmetric Learning and the Wage/Productivity Relationship," *Journal of Economic Behavior & Organization* 31(3), 419–429.

——— (forthcoming): "Theory and Evidence in Internal Labor Markets," in *The Handbook of Organizational Economics*, R. Gibbons and J. Roberts, Eds., Princeton, NJ: Princeton University Press.

TABLE 1

Percentage Distribution of Managers Across Levels by Tenure

Tenure	Separation	Level 1	Level 2	Level 3	Total
1	0.0	100.0	0.0	0.0	100.0
2	14.5	45.6	39.9	0.0	100.0
3	27.7	16.8	46.8	8.7	100.0
4	37.9	7.6	29.2	25.3	100.0
5	46.1	4.6	18.1	31.1	100.0
6	52.0	2.9	12.3	32.8	100.0
7	57.6	2.1	7.7	32.6	100.0

TABLE 2

Hazard Rates of Separation, Retention at Level, and Promotion (Percentages)

	Level 1			Level 2			Level 3	
Tenure	Separated	Retained	Promoted	Separated	Retained	Promoted	Separated	Retained
1 to 2	14.5	45.6	39.9	–	–	–	–	–
2 to 3	14.6	36.8	48.6	16.3	61.9	21.8	–	–
3 to 4	11.7	45.6	42.7	15.6	47.0	37.4	10.5	89.5
4 to 5	11.9	60.6	27.5	14.7	54.8	30.5	12.2	87.8
5 to 6	9.1	62.1	28.8	12.8	60.5	26.7	10.1	89.9
6 to 7	12.2	73.2	14.6	15.4	59.4	25.1	10.0	90.0

TABLE 3

Percentage of High Ratings at Level and Conditional on Retention at Level or Promotion

	Managers at			At Level 1		At Level 2	
Tenure	Level 1	Level 2	Level 3	Retained	Promoted	Retained	Promoted
1	52.7	–	–	52.8	55.2	–	–
2	34.8	58.4	–	30.3	37.4	54.1	74.7
3	19.6	43.9	84.1	16.9	23.5	40.9	51.5
4	11.8	26.2	54.3	10.3	19.0	24.8	35.7
5	2.4	18.7	50.0	3.6	0.0	16.2	21.7
6	3.7	12.5	43.4	0.0	25.0	11.9	16.1
7	0.0	13.0	37.1	0.0	0.0	12.1	9.1

TABLE 4
Percentage Wage Distributions by Level and Tenure

Level	Tenure	Between $20K and $40K	Between $40K and $60K	Between $60K and $80K	Mean ($)	Standard Deviation ($)
Level 1	1	59.1	40.5	0.4	39,375	6,732
	2	54.4	44.8	0.8	39,870	7,096
	3	55.6	44.4	0.0	39,552	7,234
	4	53.8	46.2	0.0	39,032	7,125
	5	64.1	35.9	0.0	37,930	6,919
	6	69.2	30.8	0.0	36,950	6,698
	7	75.0	25.0	0.0	35,613	7,250
Level 2	2	35.1	63.3	1.6	43,364	6,692
	3	31.3	65.7	2.9	43,807	7,293
	4	36.3	60.3	3.4	43,446	7,689
	5	37.1	59.6	3.3	43,340	7,898
	6	42.4	53.3	4.2	42,828	8,010
	7	41.3	55.8	2.9	42,808	7,497
Level 3	3	2.8	84.9	12.3	50,589	7,344
	4	4.5	85.3	10.1	49,882	6,999
	5	5.3	84.2	10.5	50,264	7,252
	6	6.1	84.4	9.5	50,306	7,422
	7	4.5	81.1	14.4	51,014	7,433

TABLE 5
Percentage Distribution of Changes in Log Wage by Tenure

Tenure	Between –0.15 and 0.00	Between 0.00 and 0.15	Between 0.15 and 0.30	Growth Rate
1 to 2	22.9	69.9	7.2	5.2
2 to 3	22.6	70.4	6.6	5.1
3 to 4	24.9	70.3	4.3	3.9
4 to 5	23.6	70.1	5.9	2.2
5 to 6	22.5	70.5	6.9	0.7
6 to 7	21.9	68.5	8.3	1.8

TABLE 6
Percentage Distribution of Managers Across Levels by Tenure

	Separation		Level 1		Level 2		Level 3	
Tenure	Data	Model	Data	Model	Data	Model	Data	Model
1	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0
2	14.5	14.5	45.6	45.7	39.9	39.8	0.0	0.0
3	27.7	26.5	16.8	17.2	46.8	47.3	8.7	8.9
4	37.9	37.1	7.6	8.1	29.2	29.2	25.3	25.6
5	46.1	45.3	4.6	5.3	18.1	18.3	31.1	31.2
6	52.0	51.5	2.9	3.4	12.3	12.6	32.8	32.5
7	57.6	56.9	2.1	2.7	7.7	8.3	32.6	32.1

TABLE 7
Hazard Rates of Separation, Retention at Level, and Promotion (Percentages)

		Separated		Retained		Promoted	
Level	Tenure	Data	Model	Data	Model	Data	Model
Level 1	1 to 2	14.5	14.5	45.6	45.7	39.9	39.8
	2 to 3	14.6	14.5	36.8	20.3	48.6	55.2
	3 to 4	11.7	8.4	45.6	46.9	42.7	27.8
	4 to 5	11.9	5.0	60.6	65.0	27.5	19.9
	5 to 6	9.1	5.1	62.1	64.8	28.8	21.6
	6 to 7	12.2	5.0	73.2	79.1	14.6	11.4
Level 2	2 to 3	16.3	13.6	61.9	55.5	21.8	10.9
	3 to 4	15.6	16.9	47.0	51.7	37.4	31.4
	4 to 5	14.7	14.2	54.8	56.9	30.5	28.9
	5 to 6	12.8	12.1	60.5	62.6	26.7	25.3
	6 to 7	15.4	11.5	59.4	62.9	25.1	25.6
Level 3	3 to 4	10.5	12.2	89.5	87.8		
	4 to 5	12.2	14.2	87.8	85.8		
	5 to 6	10.1	12.1	89.9	87.9		
	6 to 7	10.0	11.5	90.0	88.5		

TABLE 8
Percentage of High Ratings at Levels 1 and 2

Tenure	Level 1		Level 2	
	Data	Model	Data	Model
1	52.7	51.7	–	–
2	34.8	34.9	58.4	56.1
3	19.6	21.2	43.9	42.7
4	11.8	11.9	26.2	30.4
5	2.4	6.2	18.7	20.4
6	3.7	3.2	12.5	13.0
7	0.0	1.6	13.0	8.0

TABLE 9
Percentage Wage Distributions by Level and Tenure

Level	Tenure	Between $20K and $40K		Between $40K and $60K		Between $60K and $80K	
		Data	Model	Data	Model	Data	Model
Level 1	1	59.1	57.5	40.5	42.0	0.4	0.5
	2	54.4	57.4	44.8	41.7	0.8	0.8
	3	55.6	56.6	44.4	42.2	0.0	1.2
	4	53.8	56.7	46.2	41.9	0.0	1.4
	5	64.1	68.1	35.9	31.1	0.0	0.7
	6	69.2	69.5	30.8	29.8	0.0	0.7
	7	75.0	70.6	25.0	28.5	0.0	0.8
Level 2	2	35.1	34.3	63.3	63.8	1.6	1.8
	3	31.3	36.1	65.7	61.7	2.9	2.2
	4	36.3	36.9	60.3	60.5	3.4	2.5
	5	37.1	37.5	59.6	59.8	3.3	2.7
	6	42.4	38.1	53.3	59.0	4.2	2.9
	7	41.3	38.8	55.8	58.2	2.9	3.0
Level 3	3	2.8	8.2	84.9	82.4	12.3	9.3
	4	4.5	10.3	85.3	80.5	10.1	9.2
	5	5.3	11.3	84.2	79.1	10.5	9.5
	6	6.1	12.4	84.4	77.6	9.5	9.9
	7	4.5	13.4	81.1	76.4	14.4	10.0

TABLE 10
Estimates of Model Parameters

Parameters	Value	Asymptotic Standard Error
Prior Distribution		
ϕ_{11} (p_{11} = 0.338)	–0.672	0.022
ϕ_{21} (p_{21} = 0.381)	–0.484	0.021
ϕ_{31} (p_{31} = 0.465)	–0.141	0.017
ϕ_{41} (p_{41} = 0.607)	0.435	0.022
q_1	0.155	0.017
q_2	0.211	0.030
q_3	0.313	0.076
Probability of High Output		
α_1	0.514	0.062
β_1	0.456	0.014
α_2	0.5437	0.006
β_2	0.491	0.013
α_3	0.5435	0.007
β_3	0.490	0.010
Ratings Error		
d_0	0.521	0.040
$d_2(1)$	–0.703	0.040
$d_2(2)$	–0.544	0.029
Human Capital		
b_{124}	–704.735	3.577
b_{125}	–479.607	3.232
c_{12}	2,960.515	13.719
γ_{22} (= $c_{22} + b_{124}$)	1,858.714	1.993
γ_{23} (= $c_{23} + b_{124}$)	1,505.367	5.373
γ_{25} (= $c_{25} + b_{125}$)	1,692.309	5.156
γ_{26} (= $c_{26} + b_{125}$)	1,745.184	2.191
b_{334}	853.477	5.941
b_{335}	202.791	4.475
b_{337}	228.069	2.715
c_{31}	–399.955	9.659
c_{37}	2,190.704	1.041
c_{38}	2,003.340	1.066
Exogenous Separation		
η_{11}	0.145	0.004
ξ_3	0.033	0.001
η_{14}	0.050	0.0001
η_{21}	0.136	0.002
η_{24}	0.142	0.001
η_{25}	0.121	0.001
η_{26}	0.115	0.0003
η_{27}	0.111	0.0003
η_{31}	0.122	0.002

TABLE 10 (Continued)
Estimates of Model Parameters

Parameters	Value	Asymptotic Standard Error
Parameters of ω_{ik}(age,edu,year)		
ϖ_{011}	8.805	0.005
ϖ_{021}	9.288	0.005
ϖ_{031}	9.213	0.011
ϖ_{041}	8.865	0.013
ϖ_{012}	8.969	0.004
ϖ_{022}	9.359	0.004
ϖ_{032}	9.281	0.009
ϖ_{042}	8.945	0.012
ϖ_{013}	9.534	0.008
ϖ_{023}	9.813	0.004
ϖ_{033}	9.738	0.007
ϖ_{043}	9.418	0.011
ϖ_{1}	0.028	0.0001
ϖ_{2}	–0.0003	0.000002
ϖ_{3}	0.022	0.0004
ϖ_{13}	0.010	0.001
ϖ_{23}	–0.0001	0.00001
ϖ_{33}	0.021	0.001
ϖ_{y5}	–0.063	0.003
ϖ_{y6}	–0.107	0.004
ϖ_{y7}	–0.140	0.004
ϖ_{y8}	–0.208	0.003
ϖ_{y9}	–0.169	0.003
Coefficient on Tenure		
ω_{111}	0.007	0.0003
Coefficients on Prior by Type		
ω_{21}	2.371	0.045
ω_{22}	1.833	0.027
ω_{23}	1.316	0.015
ω_{24}	1.364	0.010
Wage Standard Deviations by Type and Level		
σ_{11}	0.076	0.001
σ_{21}	0.070	0.001
σ_{31}	0.057	0.001
σ_{41}	0.044	0.001
σ_{12}	0.063	0.001
σ_{22}	0.047	0.001
σ_{32}	0.0302	0.0004
σ_{42}	0.0303	0.0004
σ_{3}	0.047	0.0004

TABLE 11A
Counterfactual Experiments: Importance of Learning for Wages
Baseline, No Learning, Fast Learning at Level 1, Fast Learning at Level 2*

	Wages in Each Case			
			Fast Learning at	
Statistic	Baseline	No Learning	Level 1	Level 2
Means by Level				
Level 1	$39,584	$39,706	$58,271	$37,847
Level 2	43,179	43,070	61,451	77,503
Level 3	48,963	48,454	44,623	24,360
Standard Deviations by Level				
Level 1	$6,936	$6,791	$35,961	$8,668
Level 2	7,077	6,464	51,466	45,057
Level 3	8,046	6,534	45,784	4,281
Cumulative Growth Rates				
Tenure 2	4.6%	3.3%	39.3%	8.8%
Tenure 3	8.9	6.8	48.5	51.5
Tenure 4	13.8	9.8	52.8	50.1
Tenure 5	15.9	11.1	55.4	50.3
Tenure 6	17.5	12.9	58.1	50.1
Tenure 7	18.5	14.6	60.6	49.8
Tenure 7 (Balanced Panel)	19.4	15.4	62.5	51.2

*No Learning: $\beta_k = \hat{\alpha}_k, k = 1, 2, 3$; Fast Learning at Level k: α_k = 0.99 and β_k = 0.01, k = 1,2

TABLE 11B

Counterfactual Experiments: Importance of Learning for Level Assignments

Baseline, No Learning, Fast Learning at Level 1, Fast Learning at Level 2*

	Separation				Level 1				Level 2				Level 3			
Tenure	Base.	No L	Fast L at 1	Fast L at 2	Base.	No L	Fast L at 1	Fast L at 2	Base.	No L	Fast L at 1	Fast L at 2	Base.	No L	Fast L at 1	Fast L at 2
1	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	14.5	14.5	14.5	14.5	45.7	57.7	40.5	0.0	39.8	27.8	45.0	85.4	0.0	0.0	0.0	0.0
3	26.5	26.6	26.5	26.1	17.2	20.6	14.6	5.0	47.3	52.0	38.9	35.0	8.9	0.8	20.0	33.8
4	37.1	37.3	36.7	36.6	8.1	11.6	7.3	2.0	29.2	40.3	25.9	29.1	25.6	10.8	30.1	32.3
5	45.3	45.1	45.0	45.4	5.3	8.2	4.9	1.3	18.3	30.3	17.7	24.7	31.2	16.4	32.4	28.6
6	51.5	51.2	51.3	51.9	3.4	5.6	3.2	1.0	12.6	23.3	13.1	21.7	32.5	19.9	32.3	25.4
7	56.9	56.4	56.7	57.4	2.7	4.5	2.6	0.9	8.3	15.4	8.7	19.2	32.1	23.6	32.0	22.6

*No Learning: $\beta_k = \hat{\alpha}_k, k = 1,2,3$; Fast Learning at Level *k*: α_k = 0.99 and β_k = 0.01, *k* = 1,2

TABLE 12

Counterfactual Experiment: Importance of Experimentation for Level Assignments

Baseline and Equal Informativeness as Level 1*

	Separation		Level 1		Level 2		Level 3	
Tenure	Base.	Equal Info. As L1	Base.	Equal Info. As L1	Base.	Equal Info. As L1	Base.	Equal Info. As L1
1	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0
2	14.5	14.5	45.7	84.6	39.8	0.9	0.0	0.0
3	26.5	26.9	17.2	6.2	47.3	18.3	8.9	48.6
4	37.1	36.4	8.1	2.0	29.2	7.7	25.6	53.8
5	45.3	45.3	5.3	1.1	18.3	3.5	31.2	50.2
6	51.5	51.8	3.4	0.6	12.6	2.0	32.5	45.6
7	56.9	57.3	2.7	0.5	8.3	1.3	32.1	40.9

*Equal Informativeness as Level 1: $\alpha_k = \hat{\alpha}_1$, $\beta_k = \hat{\beta}_1$, $k = 2,3$