

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

Elena Pastorino

Federal Reserve Bank of Minneapolis

University of Minnesota

May 2013

Abstract

This paper develops and structurally estimates a labor market model that integrates job assignment, learning, and human capital acquisition to account for the main patterns of careers in firms. A key innovation is that the model incorporates workers' job mobility within and between firms, and the possibility that, through job assignment, firms affect the rate at which they acquire information about workers. The model is estimated using longitudinal administrative data on managers from one U.S. firm in a service industry (the data of Baker, Gibbs, and Holmström (1994a,b)) and fits the data remarkably well. The estimated model is used to assess both the direct effect of learning on wages and its indirect effect through its impact on the dynamics of job assignment. Consistent with the evidence in the literature on comparative advantage and learning, the estimated direct effect of learning on wages is found to be small. Unlike in previous work, by jointly estimating the dynamics of beliefs, jobs, and wages imposing all of the model restrictions, the impact of learning on job assignment can be uncovered and the indirect effect of learning on wages explicitly assessed. The key finding of the paper is that the indirect effect of learning on wages is substantial: overall learning accounts for one quarter of the cumulative wage growth on the job during the first seven years of tenure. Nearly all of the remaining growth is from human capital acquisition. A related novel finding is that the different speed of learning at different jobs is a primary determinant of the timing of promotions and wage increases. Along with persistent uncertainty about ability, these differences in the speed of learning across jobs are responsible for substantially compressing wage growth in low tenures.

Keywords: Careers, Job Mobility, Experimentation, Bandit, Human Capital, Wage Growth

JEL Classification: D22, D83, J24, J31, J44, J62

*I am truly indebted to Ken Wolpin for his generous advice. I benefited from numerous conversations with John Geweke and Petra Todd. Chris Ferrall, Robert Gibbons, Hiro Kasahara, Michael Keane, John Moore, and Michael Waldman have offered especially valuable comments. Finally, I thank George Baker for kindly providing me with the data and Bengt Holmström for his support of the project. 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.

The literature on careers in firms studies the allocation of workers to tasks and jobs within a firm and the dynamics of wages with tenure. Observationally, workers usually advance from lower- to higher-level jobs of a firm’s hierarchy over time, wages on average increase with tenure in a firm, especially through promotions, but wage decreases are also common, mostly in response to unsuccessful performance. (See Gibbons and Waldman (1999a,b, 2006), hereafter GW, and Waldman (2012).) Models that are qualitatively successful at explaining these patterns combine job assignment, learning about workers’ abilities, and human capital acquisition.¹ One feature that has generally been missing in the literature on careers is *experimentation* within firms: information on worker ability is assumed to be passively acquired independently of a worker’s job. The differential learning possibilities associated with different jobs within a firm can, however, have important implications for firms’ job assignment decisions. (See Prescott and Visscher (1980) and Holmström and Tirole (1989) for early references.) Another element that the literature on careers has often ignored is worker mobility between firms and the impact of competition among firms on internal job assignment and compensation. This paper is the first to provide a comprehensive examination of the empirical relevance of these models of careers, of the relative importance of their components, and of the role of these missing elements.

Specifically, this paper develops and structurally estimates a model of the labor market with experimentation and turnover that integrates elements of learning as in Jovanovic and Nyarko (1997), human capital acquisition as in Keane and Wolpin (1997), and careers in firms as in GW. The model is estimated using administrative data on managers from a single U.S. firm in a service industry (the data of Baker, Gibbs, and Holmström (1994a,b), hereafter BGH), which contain information about each manager’s yearly job assignment, salary, and performance. A key advantage of the data is that the detailed performance information identifies the process for learning at each job in the firm. The tenure profiles of managers’ job assignments and wages separately identify the process for human capital acquisition at each job. Hence, these data allow me to estimate the informativeness of each of the firm’s jobs, determine the speed of learning at each job, and assess the relative contribution of learning and human capital acquisition to wage growth on the job.

At the estimated parameters, the model implies that learning is a quantitatively important source of observed career paths, unlike common estimates in the literature on comparative advantage and learning, as in Gibbons, Katz, Lemieux, and Parent (2005), Lluís (2005), and Hunnes (2012). Conceptually, these papers measure the direct and contemporaneous dependence of wages on beliefs about ability. They do so by assessing the impact of the time-varying uncertainty about ability due to learning on the estimated effect of *current* characteristics of workers and jobs on *current* wages. This direct effect of learning on wages is estimated based on the wage process alone, instrumented to correct for the endogeneity of job assignment, and typically found to be small or insignificant or

¹That learning and human capital acquisition are primary determinants of observed earnings-experience profiles has recently been argued by Rubinstein and Weiss (2007). See Bagger, Fontaine, Postel-Vinay, and Robin (2011) for a detailed assessment of the contribution of search and human capital acquisition to individual wage growth.

difficult to reconcile with economic intuition. In sharp contrast, by explicitly estimating the joint dynamics of beliefs, job assignments, and wages implied by the model, here I am able to assess both the direct effect of learning on wages and its *indirect* effect due to its impact on the dynamics of job assignment. Intuitively, the indirect effect of learning on wages arises because learning leads managers to be more quickly promoted to higher levels of the job hierarchy, at which they are paid higher wages. Consistent with the literature, I find the direct effect of learning on wages to be small.² Critically, however, I find its indirect effect to be much larger than the direct one. I estimate that learning contributes to more than one quarter of cumulative wage growth on the job during the first seven years of tenure, with the remaining growth explained by human capital acquisition.³ This finding implies a revised view of the role of learning for wages: the impact of learning on wages is substantial, but this effect is dynamic, operating indirectly through job promotions rather than through a direct static effect. Another key finding of this paper is that the different speed of learning at different jobs, that is, the experimentation component of learning, is a primary determinant of the timing of promotions and wage increases, which leads to a compression of wages at low tenures. Without experimentation, learning would account for an even greater contribution to wage growth.

In the model, production in firms is organized among distinct *jobs* to which workers are assigned. Initially, a worker’s ability is unobserved, but over time all firms and the worker learn about ability by observing the worker’s performance. When employed, workers also acquire human capital, which can be task- and firm-specific to varying degrees. (See Sanders and Taber (2012) on the importance of task-specific human capital for wage growth.) Jobs differ in the output they generate and in the information they provide about ability. Hence, the speed of learning differs across jobs. As a consequence, when assigning a worker to a job, firms trade off current output against the value of information and future human capital. Likewise, when comparing employment at different firms, workers weigh current wages against the value of the information and human capital that can be acquired. Thus, firms and workers face a classic multi-armed bandit problem with *dependent arms*.⁴

I assume that firms compete for managers in jobs and wages in a Bertrand fashion. This formulation allows firms to be heterogeneous in their technologies and thus imperfectly competitive, but it also nests GW’s framework of firms with identical technologies. Heterogeneity in technologies generates not only wage dispersion among workers at any point in time but also worker turnover over time as workers move to firms with technologies that best match their partially learned ability and acquired human capital. With heterogeneous firms, equilibrium *endogenously* determines a flexible sharing rule of the surplus generated by a firm and a worker that does not restrict a worker to be

²By replicating the analysis of Lluís (2005) and Hunnes (2012) in my data, based on the same instrumental variable approach as in Gibbons et al. (2005), I also find that the impact of learning is negligible and insignificant. These estimates, though, do not capture the total effect of learning on wages. Details are available upon request.

³Similarly, Bagger et al. (2011) estimate that human capital acquisition accounts for 50% to 70% of the wage growth of individuals with 10 to 20 years of labor market experience.

⁴Jovanovic (1979), Miller (1984), and Flinn (1986) provide influential applications of the bandit problem with independent arms to labor market and occupational turnover.

paid his expected output at the employing firm. By measuring the difference between estimated output and wages, I can then assess the degree of monopsony power of the firm in my data.

The different role that wages play in the model, compared with their role in perfectly competitive models such as GW, is critical for my results. Under perfect competition among identical firms, the wage paid by a firm reflects a worker's value to the firm. Under Bertrand competition among heterogeneous firms, the wage reflects a worker's value to the firm's competitors. Hence, wages paid by the firm in my data provide direct information about the technologies of other firms in the labor market. As a result, certain interpretable reduced-form parameters of other firms can be recovered just by estimating the wage process at this one firm and used to assess the degree of transferability of unobserved ability and acquired human capital between the firm in my data and other firms.

The main intuition for the identification of the model is that performance data identify the process for learning *independently* of the process for human capital acquisition. Job transitions within the firm and separations identify the process for output and human capital acquisition at the firm. In particular, differences in the hazard rate of promotion across tenures and levels identify the degree of task generality of the human capital acquired at a level or with overall tenure in the firm. Finally, wages identify the degree of generality of unobserved ability and acquired human capital across firms. I estimate the model by nonparametric maximum likelihood using eight years of observations on ten cohorts of managers entering into the firm between 1970 and 1979, imposing all of the theory's restrictions. The estimated model successfully captures the tenure profile of job-to-job transitions within the firm and separations from the firm, as well as the distribution of wages and performance at the main job levels in each tenure.

The estimates of the model's parameters imply several key features 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 of the managers in my data have initial priors (that their ability is high rather than low) close to 0.5. Second, initial priors about ability are highly heterogeneous across managers, implying a significant dispersion of information at the time of hiring. Third, learning is gradual: more than 15 years of high performance are necessary for the average prior about a manager's ability being high to reach 0.9. Nonetheless, by comparing wage growth in the model and in a counterfactual scenario where belief updating is precluded, I find that learning contributes to more than 25 percent of wage growth on the job over the first seven years of tenure. The remaining growth is essentially from human capital acquisition. Intuitively, despite the direct effect of beliefs on wages being small, learning leads managers with higher ability and acquired human capital to advance more rapidly to higher levels of the job hierarchy, at which they are more productive and earn higher wages. Thus, without learning wages increase more slowly.

Interestingly, I estimate that the success rate of a manager is largest at the intermediate and highest levels of the firm's job hierarchy, but the informativeness of performance is largest at the *lowest* level. Hence, when deciding on a manager's assignment, the firm in my data experiments: it

must weigh the benefit of a high success rate against the benefit of information about ability. The firm's desire to gather information about ability leads managers to be assigned to the lowest level of the job hierarchy, at which paid wages are lowest, for a much longer period than when jobs are equally informative. As a result, without experimentation, by the fourth year of tenure managers would experience higher wage growth than their total estimated wage growth over seven years; by the seventh year of tenure, wage growth would be up to 20 percent higher.

Since learning is gradual, I find that persistent uncertainty about ability significantly reduces the pace of promotion to higher levels of the job hierarchy and hence the pace of wage growth. Absent such uncertainty, managers would be rapidly promoted to higher-level jobs and receive higher wages. Thus, *both* experimentation and uncertainty are responsible for a substantial compression of wage growth at low tenures in the firm.

Finally, I estimate that the human capital acquired with overall tenure in the firm is highly transferable across the firm's levels, unlike the human capital acquired at a level. The estimates of the wage process imply that initial human capital, as captured by a manager's education and age at entry into the firm, is highly general across firms. A comparison of the estimated output and wage processes also reveals that the firm enjoys a large degree of monopsony power.

The model shares four features with existing models of careers in firms. First, I follow Rosen (1982), Waldman (1984), and GW in allowing higher-ability workers to have a comparative advantage at higher-level jobs of a firm's hierarchy. Second, as in Jovanovic (1979), Harris and Holmström (1982), MacDonald (1982), Miller (1984), Farber and Gibbons (1996), Jovanovic and Nyarko (1997), and GW, firms and workers learn symmetrically about ability. In contrast to Jovanovic (1979), Miller (1984), and Flinn (1986), unobserved ability has a common component across jobs and firms. Third, workers can improve their productivity over time and acquire human capital that (unlike GW and in the spirit of Keane and Wolpin (1997)) can be task- and firm-specific to varying degrees. Fourth, output is linearly separable across workers.⁵

Learning about worker productivity in the labor market or in different occupations has been empirically investigated by Miller (1984), Flinn (1986), Berkovec and Stern (1991), Nagypál (2007), Groes, Kircher, and Manovskii (2010), and Papageorgiou (2012). Using a sample of executives from a French state-owned firm, Chiappori, Salanié and Valentin (1999) provide evidence of learning and downward wage rigidity in firms. Following Gibbons et al. (2005), who study sectoral and inter-industry wage differentials, Lluís (2005) and Hunnes (2012) assess the importance of comparative advantage and learning for worker mobility within firms and across occupations, respectively. None of these papers on careers, however, recover primitive parameters of uncertainty and learning at a firm's jobs or assess the contribution of learning and human capital acquisition to wage growth.

⁵For the impact of complementarities among workers on the assignment of workers to tasks or teams, see Kremer (1993), Kremer and Maskin (1996), Davis (1997), Ferrall (1997), and Ferrall, Salvanes, and Sørensen (2009).

1 A Simple Model of Careers

I start with a simplified model to develop intuition and then turn to the general model I estimate. See the Appendix and the online Supplementary Appendix for all omitted details.

Environment. A finite number of firms compete for workers over an infinite horizon with discrete dates $t \geq 1$. Firms produce a homogeneous output good, which they sell in a perfectly competitive market at a price normalized to one. Workers enter the labor market in period 1, so t also represents a worker's experience. Workers are one of two types, $\theta \in \{\alpha, \beta\}$, where α is denoted *high ability*. Firms and workers share a common initial prior belief p_1 that a worker is of high ability. Both firms and workers discount the future by the factor $\delta \in (0, 1)$.

Each firm's technology is constant returns to scale in the only input to production, labor, with workers perfectly substitutable, and consists of K^f tasks or *jobs*. Workers acquire human capital in the labor market through employment at different jobs and firms. When a worker with human capital h_t in period t is assigned to job k of firm f , output can take one of two values: high, $y_{fHk}(h_t)$, or low, $y_{fLk}(h_t)$. Realized output is observed by all firms and workers. (For analyses that relax this assumption, see Waldman (1984, 1990), Greenwald (1986), Ricart i Costa (1988), Bernhardt and Scoones (1993), Bernhardt (1995), and Waldman (1996).⁶) The probability that output is high depends on the worker's unobserved type, θ , as well as k and f , which are both observed. Denote by α_{fk} the probability a high ability worker produces high output and by β_{fk} the probability a low ability worker produces high output at job k of firm f . The type indices $\alpha = \{\alpha_{fk}\}$ and $\beta = \{\beta_{fk}\}$ are, thus, vectors of probabilities of high output at the jobs of all firms. Overall, each job is characterized by the four-tuple $\{y_{fHk}(h_t), y_{fLk}(h_t), \alpha_{fk}, \beta_{fk}\}$.

Note that human capital, h_t , affects the magnitude of the output that a worker produces but has no impact on α_{fk} or β_{fk} and, hence, no effect on the frequency of high output. This assumption will prove key to empirically distinguish learning about ability from on-the-job human capital acquisition: it allows me to identify α_{fk} and β_{fk} from the fraction of high and low performance in the data, which I assume correspond to high and low output in the model.

Since unobserved ability is a permanent characteristic of workers, it affects the quality of the match of a worker to *all* jobs and firms. Thus, learning about a worker's ability is correlated across jobs and firms.⁷ The productivity of a worker's ability, nonetheless, can differ across jobs and firms. Consider a given distribution of jobs across firms (not all firms may have all jobs) and distinguish jobs by the magnitude of high and low output. At one extreme, when α_{fk} and β_{fk} vary only across jobs, ability is job-specific but general across firms with the same jobs. At the other extreme, when α_{fk} and β_{fk} vary only across firms, ability is firm-specific but general across the jobs of each

⁶See DeVaro and Waldman (2012) for the potentially important signaling role of education in the BGH data.

⁷Relaxing the assumption of independent ability across jobs and firms, as in Jovanovic (1979), avoids the counterfactual prediction that workers change jobs within a firm or turn over only upon unsuccessful performance.

firm. The pattern of job- and firm-specificity of ability can be different for workers of high and low ability. Depending on the heterogeneity of technologies across firms, different patterns and degrees of complementarity between ability and jobs are possible. (See Pastorino (2013).)

At the beginning of t all firms and workers share a common prior p_t that a given worker is of ability α . Let $\bar{y}_f(\alpha, h_t, k) = y_{fLk}(h_t) + [y_{fHk}(h_t) - y_{fLk}(h_t)]\alpha_{fk}$ denote the expected output of a worker of ability α with human capital h_t at job k of firm f ; $\bar{y}_f(\beta, h_t, k)$ is similarly defined for a worker of ability β . Given p_t , expected output, averaged over the two possible worker types, is

$$y_f(p_t, h_t, k) \equiv \bar{y}_f(\beta, h_t, k) + [\bar{y}_f(\alpha, h_t, k) - \bar{y}_f(\beta, h_t, k)]p_t = b_{fk}(h_t) + c_{fk}(h_t)p_t, \quad (1)$$

where $b_{fk}(h_t) = \bar{y}_f(\beta, h_t, k)$ and $c_{fk}(h_t) = \bar{y}_f(\alpha, h_t, k) - \bar{y}_f(\beta, h_t, k)$ is the difference between the expected output of a worker of ability α and that of a worker of ability β with the same human capital, h_t . Note for later that $\bar{y}_f(\alpha, h_t, k) - \bar{y}_f(\beta, h_t, k) = (\alpha_{fk} - \beta_{fk})[y_{fHk}(h_t) - y_{fLk}(h_t)]$. I discuss the law of motion for h_t below. I normalize the outside option of a firm of not employing a worker to zero. I assume that the outside option of a worker of not working in the market under consideration is a sufficiently low constant, \bar{U} , that the worker chooses to work in the market in each period. I normalize payoffs by $(1 - \delta)$ so as to express them as per-period averages.

I assume that a finite number of possible technologies exist, labeled by $\{A, B, C, \dots\}$, with zero cost of adopting technologies other than A . All technologies different from A entail the same learning and human capital acquisition possibilities. (In the general model I relax these assumptions.) Note that free entry implies zero profits for any firm adopting a technology other than A . The positive cost of adopting technology A implies that at most one firm will choose it. Otherwise, since ex-post competition among firms with technology A would drive their profits to zero, such firms would make ex-ante negative profits by adopting technology A . The firm in the BGH data, referred to as *firm A* or *my firm*, is interpreted as the firm with technology A and is allowed to have incurred a strictly positive cost of adoption, that is, to be imperfectly competitive.⁸

Belief Updating. After period t production takes place, all firms and workers update beliefs about a worker's ability according to Bayes' rule, which leads to two possible values of p_{t+1} ,

$$P_{fHk}(p_t) = \frac{\alpha_{fk}p_t}{\alpha_{fk}p_t + \beta_{fk}(1 - p_t)} \text{ or } P_{fLk}(p_t) = \frac{(1 - \alpha_{fk})p_t}{(1 - \alpha_{fk})p_t + (1 - \beta_{fk})(1 - p_t)}, \quad (2)$$

depending, respectively, on whether high or low output is realized. Since the probability of high output for a worker of either ability differs across jobs, the *informativeness of a job* and the implied *speed of learning* differ across jobs. (As in Blackwell (1951), I measure the informativeness of a job by the dispersion in posterior beliefs after output is observed.)

⁸That the output market in which all firms operate, including the one in my data, is perfectly competitive whereas the labor market where they hire their workers is imperfectly competitive is consistent with an island economy with a small number of firms on each island in which output is freely mobile across islands but labor is immobile.

Timing. At the beginning of any period t , all firms simultaneously submit offers to workers, which consist of a wage and a job assignment. (These offers can be interpreted as one-sided full commitment contracts that are renegotiated period by period: as shown below, these offers are sequentially optimal in that they maximize the expected present discounted value of match surplus in each period.⁹) Next, each worker decides which offer to accept. The promised wage is then paid, production takes place, output at the accepted job at the employing firm is realized, and beliefs are updated based on observed output. (Since firms commit to the period offers they make, the timing of wage payments in a period is immaterial.) Without loss, given that workers are perfectly substitutable in production, I focus on the component game between all firms and one worker. In this game, the events in t are given by $(\mathbf{w}_t, \mathbf{k}_t, \mathbf{d}_t, z_t)$, where $(\mathbf{w}_t, \mathbf{k}_t) = \{w_{ft}, k_{ft}\}_f$ denotes the vector of each firm f 's wage and job offer; $\mathbf{d}_t = \{d_{ft}\}_f$, the vector of the worker's decisions to accept ($d_{ft} = 1$) or reject ($d_{ft} = 0$) each firm f 's offer; and z_t , an indicator for whether realized output is high or low.

Equilibrium. I restrict attention to (robust) Markov perfect equilibria for which the state variable at the beginning of any period t is $s_t = (p_t, h_t)$, which consists of the worker's prior, p_t , and acquired human capital, h_t . This is the state that firms face at the time they make their job and wage offers. The state for the worker consists of s_t and the current vector of offers, $(\mathbf{w}_t, \mathbf{k}_t)$. An equilibrium consists of offer strategies $w_{ft} = w_f(s_t)$ and $k_{ft} = k_f(s_t)$ for each firm $f \in \{1, \dots, N\}$, an acceptance strategy $\mathbf{d}_t = \mathbf{d}(s_t, \mathbf{w}_t, \mathbf{k}_t)$ for the worker with typical element $d_{ft} = d_f(s_t, \mathbf{w}_t, \mathbf{k}_t)$, $f \in \{1, \dots, N\}$, and belief updating rules $P_{fHk}(p_t)$ and $P_{fLk}(p_t)$ for all f and $k \in K^f$ such that in each period: all firms maximize the expected present discounted value of profits, the worker maximizes the expected present discounted value of wages, non-employing firms are indifferent between employing and not employing the worker at the job that maximizes their expected present discounted value of profits, and beliefs are updated as in (2).

Note that the robustness requirement of equilibrium is that losing firms make credible offers in that, if the worker deviated from the equilibrium path and accepted any of their offers, these firms would not incur a loss.¹⁰ This refinement is analogous to trembling-hand perfection and is similar to refinements used in the literature (see Bergemann and Välimäki (1996) and Felli and Harris (1996, 2004)), even in standard static Bertrand games of price competition. It implies uniqueness of equilibrium. Lifting this restriction would lead to an uninteresting multiplicity of (Markov perfect equilibrium) outcomes that, as just argued, would not be robust to small perturbations of equilibrium strategies. See the Supplementary Appendix.

⁹In terms of contracting, the model is quite different but complementary to the work of Thomas and Worrall (1988, 2007). Their work focuses on the implications of risk-sharing for wages under full information rather than on the implications of competition for wages under two-sided incomplete information, that is, learning, as in my case. For example, in their limited commitment framework, if workers were risk neutral, as they are in my model, then equilibrium would imply static wage contracts very different from mine, which involve dynamic considerations.

¹⁰This requirement rules out offers higher than the present value of output (net of the continuation profit after the worker is employed by the winning firm) on the part of losing firms. Offers smaller than this present value are already ruled out by equilibrium: losing firms must prefer not employing the worker to employing him in equilibrium.

Given the strategies of all firms, at each state the worker's strategy satisfies the Bellman equation

$$V^w(s_t, \mathbf{w}_t, \mathbf{k}_t) = \max_{\mathbf{d}} \sum_f d_f [(1 - \delta)w_{ft} + \delta EV^w(s_{t+1}, \mathbf{w}_{t+1}, \mathbf{k}_{t+1} | s_t, k_{ft})], \quad (3)$$

where $\mathbf{d} = (d_1, \dots, d_N)$ denotes the collection of the worker's possible acceptance ($d_f = 1$) or rejection ($d_f = 0$) decisions with respect to each firm f 's offer of the wage $w_{ft} = w_f(s_t)$ and the job $k_{ft} = k_f(s_t)$ and $\sum_f d_f = 1$. Recall that payoffs are normalized by $(1 - \delta)$ so as to be expressed as per-period averages. The expectation EV^w is over the worker's output realized at the end of the period, so conditional on the worker's assignment to job k of firm f , the beginning-of-period state in $t + 1$ is $s_{t+1} = (P_{fHk}(p_t), h_{t+1})$ after high output and $s_{t+1} = (P_{fLk}(p_t), h_{t+1})$ after low output at job k of firm f . Note from (3) that when evaluating the prospect of employment at different firms, a worker compares the wage he is offered to the value of the information (and future human capital) that will be acquired at the proposed job, which affects future wages.

Given the strategies of the other firms and the worker, at each state the strategy of any firm f solves the Bellman equation

$$\Pi^f(s_t) = \max_{w,k} \left(d_{ft} \{ (1 - \delta)[y_f(s_t, k) - w] + \delta E\Pi^f(s_{t+1} | s_t, k) \} + \sum_{g \neq f} d_{gt} \delta E\Pi^f(s_{t+1} | s_t, k_{gt}) \right), \quad (4)$$

where $d_{ft} = d_f(s_t, \mathbf{w}_t, \mathbf{k}_t)$ and $d_{gt} = d_g(s_t, \mathbf{w}_t, \mathbf{k}_t)$ are the acceptance decisions of the worker with respect to firm f 's and firm g 's offers, respectively, $k_{gt} = k_g(s_t)$ is the job offered by firm g , and $E\Pi^f$ is the expectation over the worker's output realized at the end of the period, conditional on the assigned job: job k if firm f employs the worker or job k_{gt} if firm g employs the worker. Observe from (4) that in choosing a job for the worker, a firm trades off the worker's current expected output in a job against the value of the information revealed by output (in addition to the value of the acquired human capital), which influences future profits. Note also from the last term in (4) that in deciding which wage and job offer to make, a firm takes into account the option value of not employing the worker in the current period and attracting him in some future period. This option value is associated with observing the worker's performance at competing firms and with the human capital that the worker can acquire at these firms. In particular, both workers and firms *experiment*, that is, contemplate sacrificing current wages and output, respectively, to acquire information about ability.

If the sets of jobs $\{y_{fHk}(h_t), y_{fLk}(h_t), \alpha_{fk}, \beta_{fk}\}$ were identical across firms, then the labor market would be perfectly competitive, workers would be paid their expected output at the employing firm, and no turnover would occur (as long as there is an infinitesimal cost of switching firms). Differences in α_{fk} and β_{fk} across jobs would still lead to experimentation within firms. If α_{fk} and β_{fk} were also identical across jobs, then no experimentation would take place within firms either. In this case, the problems of the worker and the firms would be static. This special case without experimentation and without turnover is the one studied by GW.

Job Assignment. For simplicity, denote the worker's value at state s_t by $V^w(s_t)$ using the fact that in equilibrium the wage and job offers of all firms, $\mathbf{w}_t(s_t)$ and $\mathbf{k}_t(s_t)$, depend on s_t . From now on I will focus on the behavior of firm A . Denote by $V^A = \Pi^A + V^w$ the *match surplus value* of firm A , which is the sum of the (expected present discounted) value of profits of firm A (see (4) with $f = A$) and wages of the worker in (3). Denote by $V^A(s_t, f)$ the match surplus value of firm A when the worker accepts the offer of firm $f \in \{A, B, C, \dots\}$ at state s_t and by $f_{0t} = f_0(s_t)$ the *best competitor* of firm A , which is the competitor of firm A most preferred by the worker. When firm A employs the worker, firm f_{0t} is the second-best firm, that is, the firm offering the second-highest value of wages. When firm A does not employ the worker, firm f_{0t} is the best firm, that is, the employing firm. Denote by $k_{0t} = k_0(s_t)$ the job offered by firm f_{0t} .

When deciding whether to employ the worker and at which job, it can be shown by Bertrand logic that firm A solves a pseudo-planning problem with value $V^A(s_t)$ given by

$$\max \left\{ \max_{k \in K^A} \{ (1 - \delta)y_A(s_t, k) + \delta EV^A(s_{t+1}|s_t, k) \}, \max_{f \neq A, k \in K^f} \{ (1 - \delta)y_f(s_t, k) + \delta EV^A(s_{t+1}|s_t, k) \} \right\} \quad (5)$$

that is, $V^A(s_t) = \max\{V^A(s_t, A), V^A(s_t, f_{0t})\}$. I will discuss the logic behind (5) in the context of the example in Section 1.1 and in the general model. Briefly, $V^A(s_t, A)$, the value of the first inner maximization in (5), is the value to firm A and the worker of the worker's employment at firm A . $V^A(s_t, f_{0t})$, the value of the second inner maximization in (5), is the value to firm A and the worker of the worker's employment at firm f_{0t} . Conditional on employing the worker, firm A chooses the job that yields that highest match surplus value, $V^A(s_t, A)$. Firm and worker optimality also imply that in equilibrium firm A employs the worker if the associated match surplus value, $V^A(s_t, A)$, is higher than the one associated with the worker's employment at firm f_{0t} , $V^A(s_t, f_{0t})$.

To see how $V^A(s_t, f_{0t})$ is determined, recall that firm A 's competitors make zero profits. Thus, the worker extracts all the surplus from employment at any such firm, in which case his value is

$$\max_{f \neq A, k \in K^f} \{ (1 - \delta)y_f(s_t, k) + \delta EV^w(s_{t+1}|s_t, k) \}, \quad (6)$$

where $V^w = V^f$ since $\Pi^f = 0$. Now, given that the jobs of firm A 's competitors entail the same prospects for information and human capital acquisition, $E\Pi^A(\cdot|s_t, k)$ is independent of the job k chosen by a competitor. In turn, $E\Pi^A(\cdot|s_t, k_{0t}) = E\Pi^A(\cdot|s_t, k)$ for any job k of firm $f \neq A$ implies that the sum of (6) and $\delta E\Pi^A(\cdot|s_t, k_{0t})$, which is $V^A(s_t, f_{0t})$ by definition, gives the same value as the maximization of $(1 - \delta)y_f(s_t, k) + \delta[EV^w(\cdot|s_t, k) + E\Pi^A(\cdot|s_t, k)]$. The maximized value of this latter expression, since $EV^A = EV^w + E\Pi^A$, is precisely $V^A(s_t, f_{0t})$ in (5). Therefore, the states at which firm A employs the worker in equilibrium and the assigned job at firm A can be determined just by solving (5).

Wages. The wage that firm A pays also follows from Bertrand logic. Denote by $y_{f_{0t}}(s_t, k_{0t})$ the

worker's expected output at firm f_{0t} and by V^0 the match surplus value of firm f_{0t} . In equilibrium, when the worker is employed by firm A , the worker must be indifferent between the offers by firms A and f_{0t} . By the definition of equilibrium, firm f_{0t} is indifferent between employing and not employing the worker. These observations imply that, when employed by A , the worker is paid his expected output at firm f_{0t} at its offered job, k_{0t} , plus a *compensating wage differential* that reflects any difference in informativeness and prospect for human capital acquisition between jobs k_{0t} and k_{At} . Formally, the worker's wage when employed by firm A at job k_{At} is

$$w_A(s_t, \varepsilon_t) = y_{f_{0t}}(s_t, k_{0t}) + \Psi(s_t, k_{At}), \quad (7)$$

where $\Psi(s_t, k_{At}) = \delta [EV^0(\cdot|s_t, k_{0t}) - EV^0(\cdot|s_t, k_{At})] / (1 - \delta)$ is the compensating wage differential; see the general model for details. As in search models with competitive poaching like Bagger et al. (2011), competition implies that the worker is paid a varying fraction of the surplus value generated with the employing firm, unless the poacher, here firm f_{0t} , has the same technology as the employing firm. In this latter case, the worker is paid his expected output at firm A in each period, thus capturing the entire surplus of the match with firm A .

A nontrivial sharing rule of the surplus between firm A and the worker emerges here endogenously both through the dependence of current wages on expected output at other firms, $y_{f_{0t}}(s_t, k_{0t})$, and through $\Psi(s_t, k_{At})$. In particular, through $\Psi(s_t, k_{At})$ the worker extracts any difference in the continuation surplus of firm f_{0t} if the worker was employed by firm f_{0t} rather than by firm A in the current period. Importantly, all of this difference is priced in the current wage through $\Psi(s_t, k_{At})$. In contrast, in Bagger et al. (2011), a worker captures, through a given bargaining weight, a fraction of the difference in continuation surplus between a continued match with the employing firm and a match with the poacher. This difference affects the worker's present value of wages as well as current wage, through revisions of the contractual share of the output at the employing firm that the worker receives as wage. By construction, then, in Bagger et al. (2011) the worker can be paid at most his current output in a period. Here, on the contrary, the wage paid by firm A can be smaller or *larger* than a worker's expected output at firm A , depending, for instance, on the relative informativeness of job k_{0t} of firm f_{0t} and job k_{At} of firm A . In the following example, I illustrate these features of the model and its identification.

1.1 An Example

Suppose the market consists of firm A and two firms of type B and the time horizon is two periods. I begin with a simple setting without human capital acquisition in which firm A has two jobs, referred to as $A1$ and $A2$, and each firm of type B has only one job, referred to as $B1$. This setting is sufficient to show that equilibrium implies a nondegenerate and imperfectly assortative distribution of workers to jobs and firms and naturally generates job-to-job mobility between firms

in equilibrium, as well as wage increases (and possibly promotion) in response to good performance and wage decreases (and possibly demotion) in response to bad performance. In addition, wages paid by firm A do *not* equal a worker's expected output at firm A . In this setting, however, the compensating wage differential is zero. I then show that a compensating wage differential arises when each firm B has a second job or in the presence of human capital acquisition.

For simplicity, assume that job $A1$ is uninformative about ability ($\alpha_{A1} = \beta_{A1}$), job $A2$ is moderately informative ($\alpha_{A2} = \alpha > \beta_{A2} = \beta$ and $\alpha, \beta \in (0, 1)$), and job $B1$ is perfectly informative ($\alpha_{B1} = 1$ and $\beta_{B1} = 0$). Then, the only nontrivial updating rules are for job $A2$, which I simplify from $P_{AH2}(p)$ and $P_{AL2}(p)$ to $P_H(p)$ and $P_L(p)$. The initial prior belief that a worker is of high ability is simply p . The worker's expected output at firm $f \in \{A, B\}$ in job k in any period is then $y_f(p, k) = b_{fk} + c_{fk}p$. As depicted in Figure 1, I assume that ability and jobs are complementary: a worker known to be of low ability is best suited to $A1$, next-best suited to $A2$, and least suited to $B1$ whereas a worker known to be of high ability is best suited to $B1$, next-best suited to $A2$, and least suited to $A1$. Hence, in the absence of uncertainty, matching would be perfectly assortative with high ability workers employed by firms B and low ability workers employed by firm A at job $A1$. (In estimation, I do not impose any form of vertical sorting.)

Job Assignment

In period 2, job assignment is a static decision. Define p_{A2}^s as the static cutoff prior between jobs $A1$ and $A2$, which satisfies $y_A(p_{A2}^s, 1) = y_A(p_{A2}^s, 2)$. Similarly, define p_{B1}^s as the static cutoff prior between jobs $A2$ and $B1$, which satisfies $y_A(p_{B1}^s, 2) = y_B(p_{B1}^s, 1)$. Note that due to Bertrand competition, type B firms behave as perfect competitors and pay a worker his expected output in each period. Let V_2^A denote firm A 's match surplus value in the second period. Now, when firm A employs the worker, V_2^A is simply the worker's expected output at firm A at the assigned job, $y_A(p, k)$. When a firm of type B employs the worker, V_2^A is just the worker's wage (firm A 's profits are zero), which equals the worker's expected output at a firm of type B . Hence, firm A 's match surplus value in period 2 is $V_2^A(p) = \max\{\max_k y_A(p, k), y_B(p, 1)\}$. By complementarity between ability and jobs, the solution to firm A 's match surplus value problem implies that job $A1$ is assigned at $p < p_{A2}^s$, job $A2$ at $p \in [p_{A2}^s, p_{B1}^s)$, and job $B1$ at $p \geq p_{B1}^s$. It is easy to see that this is also the equilibrium assignment rule.

In period 1 the match surplus value of firm A is $V_1^A(p) = \max\{\max_k V_1^A(p, k), V_1^A(p, B)\}$, where

$$V_1^A(p, k) = (1 - \delta)y_A(p, k) + \delta\{r_{Ak}(p)V_2^A(P_{AHk}(p)) + [1 - r_{Ak}(p)]V_2^A(P_{ALk}(p))\},$$

$k = 1, 2$, by logic similar to the one in period 2. Here, $r_{Ak}(p) = \alpha_{Ak}p + \beta_{Ak}(1 - p)$ is the probability of high output at job k , $V_2^A(P_{AHk}(p))$ is the continuation match surplus value after high output, and $V_2^A(P_{ALk}(p))$ after low output. $V_1^A(p, B)$ is similarly defined: the one-period match surplus value is $y_B(p, 1)$, since firm A 's profits are zero, and the probability of high output is $r_{B1}(p) = p$.

Since job $A2$ is more informative than job $A1$, the cutoff prior p_{A2}^d at which firm A is indifferent between jobs $A1$ and $A2$ in the first period satisfies

$$p_{A2}^d < p_{A2}^s. \quad (8)$$

Hence, in equilibrium the worker is assigned to job $A2$ in the first period at *lower* priors than the lowest priors at which he is assigned to $A2$ in the second period. In particular, at priors between p_{A2}^d and p_{A2}^s , a worker is assigned to job $A2$ even if his output would be higher at job $A1$. In this sense, firm A and the worker *experiment*: they sacrifice (expected) output and wages in the current period in order to acquire information for the benefit of output and wages in the next period. Likewise, since job $B1$ is more informative than $A2$, the cutoff prior p_{B1}^d at which firm A is indifferent between employing the worker at job $A2$ and losing him to a firm of type B in the first period satisfies

$$p_{B1}^d < p_{B1}^s. \quad (9)$$

The solution to firm A 's match surplus value problem in the first period implies that job $A1$ is assigned at $p < p_{A2}^d$, job $A2$ at $p \in [p_{A2}^d, p_{B1}^d)$, and job $B1$ at $p \geq p_{B1}^d$. Again, it is easy to verify that this is also the equilibrium assignment rule in the first period. Thus, the states at which firm A employs the worker in equilibrium and its job choice can be determined simply by solving firm A 's match surplus value problem.

For concreteness, consider the outcomes depicted in Figure 2 for workers who start in job $A2$. A worker with prior p_{A2}^d stays in job $A2$ after a success and a worker with the highest initial prior at job $A2$ (just below p_{B1}^d) stays in $A2$ after a failure. Let P_H^{-1} and P_L^{-1} denote the inverse functions of P_H and P_L , respectively. Panel A in Figure 3 illustrates the outcome after a success (high output) under these assumptions. Workers with p below $P_H^{-1}(p_{B1}^s)$ stay in job $A2$ since $P_H(p) < p_{B1}^s$ and workers with p above it move to a firm of type B since $P_H(p) \geq p_{B1}^s$. Panel B in Figure 3 illustrates the outcome after a failure (low output). Workers with p below $P_L^{-1}(p_{A2}^s)$ are demoted to $A1$ since $P_L(p) < p_{A2}^s$ and workers with p above it stay in job $A2$ since $P_L(p) \geq p_{A2}^s$.

Wages

As discussed, each firm of type B pays the worker his expected output in each period. In the second period, Bertrand competition implies that the wage paid by firm A is the worker's expected output at a firm of type B . Consider now the first period wage paid by firm A to a worker employed at job $A2$. In equilibrium, the worker must be indifferent between working at job $A2$ and at a firm of type B . Based on the outcomes characterized above, a worker employed at job $A2$ after a success is either retained at job $A2$ or employed by a firm of type B . In both cases, he is paid $y_B(P_H(p), 1)$. After a failure at job $A2$, the worker is either retained at job $A2$ or demoted to job $A1$. In both cases, he is paid $y_B(P_L(p), 1)$. Since by (1) expected output is affine in the prior and

the mean of the posterior is the prior, the present value of wages for a worker initially in $A2$ is then $(1 - \delta)w_A(p) + \delta y_B(p, 1)$.

Now, a worker initially employed by a firm of type B is paid $y_B(p, 1)$ in the first period. After a success at job $B1$, the worker is retained and paid $y_B(1, 1)$. After a failure at job $B1$, the worker moves to job $A1$ of firm A and is paid $y_B(0, 1)$. So, the present value of wages for a worker initially in $B1$ is $y_B(p, 1)$. Equating the two present values of wages at jobs $A2$ and $B1$ yields that the wage paid by firm A in the first period to a worker in job $A2$ is $w_A(p) = y_B(p, 1)$, which is smaller than the worker's expected output at firm A , $y_A(p, 2)$, since $p < p_{B1}^s$. Yet, in this case

$$\Psi(p, 2) \equiv \frac{\delta}{1 - \delta} [EV_2^B(p'|p, k_{B1}) - EV_2^B(p'|p, k_{A2})] = 0, \quad (10)$$

given that, as just argued, $EV_2^B(p'|p, k_{B1}) = EV_2^B(p'|p, k_{A2}) = y_B(p, 1)$ regardless of the worker's first period assignment. Intuitively, the extra information $B1$ provides relative to $A2$ has no impact on the future output of a firm of type B when such a firm has only one job, which implies that the compensating wage differential is zero. Indeed, the extent to which a worker captures the value of the information generated at firm A depends on the impact of this information on the match surplus value of a firm of type B , as shown next.

Wages and Compensating Differential

To see how a nonzero compensating differential arises, suppose that the type B firms have two jobs and their best assignment is job $B1$ at $p = 0$ and job $B2$ at $p = 1$. Let $B1$ and $B2$ be perfectly informative. Denote by p_{B2} the static and dynamic cutoff prior such that $y_B(p, 1) = y_B(p, 2)$. Suppose that $B1$ is dominated by job $A1$ at all priors and that $p_{A2}^s < p_{B2}$. Consider workers with priors between p_{A2}^d and p_{B1}^d such that $P_H(p) < p_{B2}$, so their best assignment at a firm of type B is $B1$ in both periods. For these workers to be indifferent between jobs $A2$ and $B1$, the first period wage in job $A2$ must be $w_A(p) = y_B(p, 1) + \Psi(p, 2)$ with

$$\Psi(p, 2) = \frac{\delta}{1 - \delta} [py_B(1, 2) + (1 - p)y_B(0, 1) - y_B(p, 1)] > 0,$$

where $py_B(1, 2) + (1 - p)y_B(0, 1)$ is EV_2^B after job $B1$ is assigned in the first period and $y_B(p, 1)$ is EV_2^B after job $A2$ is assigned in the first period. That $\Psi(p, 2) > 0$ follows by second-order stochastic dominance: V_2^B is convex in p and job $B1$ is more informative than job $A2$, so the posterior beliefs reached when job $B1$ is assigned are a mean-preserving spread of the posteriors reached when job $A2$ is assigned. In this case, the extra information $B1$ provides relative to $A2$ has an impact on the future output of a firm of type B . Hence, this information has value to a firm of type B that is priced in equilibrium in the wage paid by firm A .

Wages and Expected Output

Even when the compensating wage differential is zero, in general the wage paid by firm A is

different from a worker's expected output. As shown above, when a firm of type B has only one job, a worker employed by firm A at job $A2$ in the first period is paid $w_A(p) = y_B(p, 1) < y_A(p, 2)$ since $p < p_{B1}^s$. To see how wages *larger* than expected output arise, suppose now that job $A2$ is *more* informative than job $B1$ and again that each firm of type B has only one job. By logic analogous to the above one, it follows that $p_{B1}^s < p_{B1}^d$ due to the informational advantage of job $A2$ over job $B1$. Thus, for workers in job $A2$ with priors above p_{B1}^s and below p_{B1}^d , the wage in the first period is $w_A(p) = y_B(p, 1) > y_A(p, 2)$ since $p > p_{B1}^s$. That is, firm A pays the worker *more* than his expected output in order to acquire information about his ability.

Human Capital

Suppose workers can acquire human capital when employed. In this case, the compensating wage differential need not be zero even if a firm of type B has only one job as long as the accumulation process differs across firms. To see this, suppose human capital acquisition takes the form $b_{fk}(h_t) = b_{fk} + h_{fk}(k_{t-1}) \cdot (t - 1)$ and $c_{fk}(h_t) = c_{fk} + h_{fk}(k_{t-1}) \cdot (t - 1)$, $f = A, B$; this formulation is similar to the one in the general model I estimate. Then, expected output depends on the job the worker has been assigned to in the previous period, k_{t-1} , and the worker's labor market experience, $t - 1$. The extent to which human capital is transferable across jobs and firms depends on $h_{fk}(k_{t-1})$. It can be shown that the first period wage of workers in job $A2$ is now

$$w_A(p) = y_B(p, 1) + \frac{\delta}{1 - \delta} (b_{B1} + h_{B1}(B1) + [c_{B1} + h_{B1}(B1)]p - \{b_{B1} + h_{B1}(A2) + [c_{B1} + h_{B1}(A2)]p\})$$

where $\Psi(p, 2)$, which is the second term of $w_A(p)$, is positive if, and only if, the human capital acquired at job $A2$ of firm A is *less* valuable to firms of type B than the human capital acquired at job $B1$, that is, $h_{B1}(B1) > h_{B1}(A2)$. Thus, the wage compensates the worker for forgoing the acquisition of human capital at a firm of type B that is more valuable (to the match between such a firm and the worker) than the human capital acquired at job $A2$. It is also possible that a worker is never demoted to $A1$ after low output, as commonly observed: the increase in output at jobs $A2$ and $B1$ due to the human capital acquired at job $A2$ can offset the adverse belief revision.

Identification

Consider the more general case in which both jobs $A1$ and $A2$ are informative about ability and human capital acquisition is present. The main intuition for the separate identification of the processes for learning and human capital acquisition at firm A is simple. In the model human capital accrues deterministically with experience, as in a standard learning-by-doing setup, and it only affects a worker's output at the job k of each firm f , $y_{fHk}(h_t)$ or $y_{fLk}(h_t)$. In particular, human capital has no impact on the parameters of learning, α_{fk} and β_{fk} . (Technically, the law of motion for the unobserved state, that is, Bayes' rule, is known up to the parameters α_{fk} and β_{fk} .) The Supplementary Appendix shows that in this framework, repeated performance information at each job of firm A is sufficient to identify p and $\{\alpha_{Ak}, \beta_{Ak}\}_{k=1}^2$. To see why, suppose that p is

known. Consider individuals assigned to job $A1$ in both periods and equate the sample proportions of individuals with a success at the end of periods 1 and 2, denoted, respectively, by m_{A11}^o and $m_{A12|H1}^o$, to their theoretical counterparts, $r_{A1}(p)$ and $r_{A1}(P_{AH1}(p))$. Then,

$$\begin{cases} \alpha_{A1}p + \beta_{A1}(1-p) = m_{A11}^o \\ \frac{\alpha_{A1}^2 p}{\alpha_{A1}p + \beta_{A1}(1-p)} + \frac{\beta_{A1}^2 (1-p)}{\alpha_{A1}p + \beta_{A1}(1-p)} = m_{A12|H1}^o \end{cases}$$

by using the definition of $P_{AH1}(p)$. This system has a unique solution for α_{A1} and β_{A1} , so α_{A1} and β_{A1} are identified. Three periods of repeated observations on performance at job $A1$ are sufficient to identify p as well. Similarly, two periods of observations of performance at $A2$ identify α_{A2} and β_{A2} . An analogous argument holds in the presence of classification error.

Next, the output and human capital process at firm A is identified from data on job assignments alone by the argument in Magnac and Thesmar (2002) assuming δ is known. Under standard assumptions, Magnac and Thesmar prove that dynamic discrete choice models are nonparametrically identified up to the continuation value from one alternative, K . Note that the problem of firm A consists of the choice among job $A1$, $A2$, and ‘employment at a firm of type B ’. Hence, by interpreting alternative K as employment at a firm of type B and parameterizing $EV^A(p_t, t, B)$ as a flexible polynomial, the differences between b_{Ak} , c_{Ak} , and $h_{Ak}(\cdot)$ and the corresponding parameters at the second-best firm, a type B firm, are identified. See the Supplementary Appendix for details.

2 A General Model of Careers

Now assume that all firms operate exclusive technologies and, thus, are imperfectly competitive. In this setting I also introduce productivity shocks, realized at the beginning of each period, and separation shocks, realized (if at all) at the end of a period.¹¹

Output in period t is now either $y_{fHk}(h_t) + \varepsilon_{fkt}$ or $y_{fLk}(h_t) + \varepsilon_{fkt}$, where ε_{fkt} is a mean-zero type I extreme value productivity shock with c.d.f. G , which captures *idiosyncratic* features of the match between a worker and a job, and is independent across jobs, time periods, and of the worker’s ability. In contrast, as before, a worker’s ability θ captures *permanent* features of the match between a worker and a job. The expected output of a worker with prior p_t and human capital h_t at job k of firm f after productivity shocks are realized, but before $y_{fHk}(h_t)$ or $y_{fLk}(h_t)$ is realized, is then $y_f(p_t, h_t, k) + \varepsilon_{fkt}$. See the Appendix for a discussion of this formulation of output.

To allow for separations unrelated to ability or performance, I assume that after production a worker leaves the market with probability $1 - \eta_{fk}(h_t)$. Hence, a job is now characterized by the six-tuple $\{y_{fHk}(h_t), y_{fLk}(h_t), \alpha_{fk}, \beta_{fk}, \varepsilon_{fkt}, \eta_{fk}(h_t)\}$.¹² The state that firms face at the time they

¹¹See Ghosh (2007) for a model of learning, human capital acquisition, and competition among homogeneous firms, in which turnover stems from disutility shocks to continuing employment with the same firm.

¹²As long as new workers enter into the market each period, the market will always be populated by workers.

make their job and wage offers consists of s_t and the vector of productivity shocks realized at the jobs of each firm, ε_t . The state that the worker faces when choosing which offer to accept consists of s_t , the productivity shocks of all firms, ε_t , and the current vector of offers of all firms, $(\mathbf{w}_t, \mathbf{k}_t)$.

As before, by combining the optimality conditions of firm A and the worker, it is possible to show that the best-response problem determining the job choice of firm A and the set of states at which firm A employs the worker corresponds to the problem of maximizing the match surplus value of firm A , V^A . V^A solves the Bellman equation in the next proposition. As before, I will denote the worker's value compactly by $V^w(s_t, \varepsilon_t)$, the best competitor of firm A at state (s_t, ε_t) by $f_{0t} = f_0(s_t, \varepsilon_t)$, its wage offer by $w_0(s_t, \varepsilon_t)$, its job offer by $k_0(s_t, \varepsilon_t)$, and the associated probability of exogenous separation by $1 - \eta_{0t}(h_t)$.

Proposition 1. *Firm A 's match surplus value, $V^A(s_t, \varepsilon_t) = \max\{V^A(s_t, \varepsilon_t, A), V^A(s_t, \varepsilon_t, f_{0t})\}$, is*

$$\max \left\{ \max_{k \in K^A} \left\{ (1 - \delta)[y_A(s_t, k) + \varepsilon_{Akt}] + \delta \eta_{Ak}(h_t) \int_{\varepsilon_{t+1}} EV^A(s_{t+1}, \varepsilon_{t+1} | s_t, k) dG \right\}, \right. \\ \left. (1 - \delta)w_0(s_t, \varepsilon_t) + \delta \eta_{0t}(h_t) \int_{\varepsilon_{t+1}} EV^A(s_{t+1}, \varepsilon_{t+1} | s_t, k_{0t}) dG \right\}. \quad (11)$$

To interpret (11), consider first the inner maximization problem, whose value is $V^A(s_t, \varepsilon_t, A)$. This problem characterizes the job choice of firm A if it employs the worker at state (s_t, ε_t) . If firm A employs the worker, then the current period value, $y_A(s_t, k) + \varepsilon_{Akt}$, consists of the sum of firm A 's current expected profits at the chosen job k , $y_A(s_t, k) + \varepsilon_{Akt} - w$, and the worker's wage, w . The future value, $EV^A(\cdot | s_t, k)$, is the sum of the continuation value of firm A 's profits and of the worker's wages, conditional on the assignment of job k . This inner problem implies that when firm A employs the worker, it assigns the worker to the job that maximizes match surplus.

Consider next the outer maximization problem in (11), which involves a comparison between $V^A(s_t, \varepsilon_t, A)$ and $V^A(s_t, \varepsilon_t, f_{0t})$. The term $V^A(s_t, \varepsilon_t, f_{0t})$ in the second line of (11) is firm A 's match surplus value when the worker is employed by a firm different from A . This firm is f_{0t} by definition: in equilibrium when firm A employs the worker, f_{0t} is the second-best firm, when firm A does not employ the worker, firm f_{0t} is the employing firm. If the worker chooses to work for firm f_{0t} , the current period match surplus value of firm A is just the worker's wage at firm f_{0t} , $w_0(s_t, \varepsilon_t)$, whereas the future value is $EV^A(\cdot | s_t, k_{0t})$, which is the sum of the continuation value of firm A 's profits and of the worker's wages, conditional on the assignment of job $k_{0t} = k_0(s_t, \varepsilon_t)$. This outer problem is an immediate implication of firm and worker optimality. When the worker is employed by firm A , both firm A and the worker prefer this outcome to the one in which the worker is employed by firm f_{0t} , taking into account the value of the information and human capital that can be acquired at firm f_{0t} . When the worker is employed by firm f_{0t} , both firm A and the worker prefer employment

Given the assumed separability of workers in production, this aspect is irrelevant for my analysis.

at firm f_{0t} to employment at firm A . By summing firm A 's and the worker's values in each case, it follows that $V^A(s_t, \varepsilon_t)$ must achieve the maximum between $V^A(s_t, \varepsilon_t, A)$ and $V^A(s_t, \varepsilon_t, f_{0t})$.

Proposition 1 has two crucial implications. First, the match surplus value problem determines not only the job chosen by firm A but also the set of states at which firm A employs the worker, that is, the states (s_t, ε_t) at which $V^A(s_t, \varepsilon_t, A)$ is larger than $V^A(s_t, \varepsilon_t, f_{0t})$. Second, the match surplus value problem is analogous to a pseudo-planning problem among $K^A + 1$ alternatives with the difference that here firm A has monopsony power and is competing against other firms. Hence, the alternative to employment at one of the K^A jobs of firm A is determined by equilibrium and corresponds to employment at a firm different from A .

Proposition 1 applies to any market structure. In estimation, due to data limitations, I focus on three market structures: perfect competition, duopoly, and an oligopoly in which all competitors of firm A face no cost of technology adoption and their technologies entail the same prospect for learning, human capital acquisition, and exogenous separation. This type of oligopoly is similar to the case of a dominant firm and a competitive fringe. The perfectly competitive case is a special case in which all firms have identical technologies, so $V^A(s_t, \varepsilon_t, f_{0t}) = V^A(s_t, \varepsilon_t, A)$. When the labor market is a duopoly between firms A and B , then $V^A(s_t, \varepsilon_t, f_{0t}) = V^A(s_t, \varepsilon_t, B)$ and $V^A(s_t, \varepsilon_t) = V^A(s_t, \varepsilon_t, A)$, where $V^A(s_t, \varepsilon_t, A)$ is the value of the inner maximization problem in (11). This result that $V^A(s_t, \varepsilon_t) = V^A(s_t, \varepsilon_t, A)$ is due to the equilibrium 'double indifference' of the worker between the offers by A and B and of the non-employing firm between employing and not employing the worker. Lastly, when the labor market is an oligopoly as described,

$$V^A(s_t, \varepsilon_t, f_{0t}) = \ln \sum_{f \neq A, k \in K^f} \exp \left\{ (1 - \delta) y_f(s_t, k) + \delta \eta_{fk}(h_t) \int_{\varepsilon_{t+1}} EV^A(\cdot | s_t, k) dG \right\} + (1 - \delta) \varepsilon_{0t}, \quad (12)$$

where ε_{0t} is the productivity shock realized at the job offered by the best competitor of firm A .¹³ These three market structures are nested by (11) with $V^A(s_t, \varepsilon_t, f_{0t})$ as in (12) and imply a convenient simplification of (11). For all other market structures, the formulation of the employment and job assignment problem I use in estimation is an approximation of the exact one. See the Appendix.

In the perfectly competitive case, equilibrium is efficient and firm A solves the social planner's problem. In the duopoly and oligopoly cases, equilibrium assignment does *not* necessarily solve the social planner's problem: firms do not take into account their competitors' profits when choosing their jobs. Further, wages paid by firm A do *not* equal expected output at firm A . Yet, in all these cases, the states at which firm A employs the worker and the job it chooses can be determined by solving the match surplus value problem in (11). The value functions of Proposition 1 provide the first set of estimating equations, which yield the probability of employment at firm A and of

¹³Without loss, if any two firms or more have identical technologies and, thus, correlated productivity shocks, only one is considered. Any such firm behaves as a perfect competitor, since its profits are zero.

assignment to one of its jobs. The next result derives the expression for paid wages, which provides the second main estimating equation.

Proposition 2. *The worker's equilibrium wage when employed by firm A at job k_{At} is*

$$w_A(s_t, \varepsilon_t) = y_{f_{0t}}(s_t, k_{0t}) + \Psi(s_t, k_{At}) + \varepsilon_{0t} \quad (13)$$

where $\Psi(s_t, k_{At}) = \frac{\delta}{1-\delta} \int_{\varepsilon_{t+1}} [\eta_0(h_t)EV^0(\cdot|s_t, k_{0t}) - \eta_{Ak_{At}}(h_t)EV^0(\cdot|s_t, k_{At})] dG$.

See the Supplementary Appendix for the proof. By Proposition 2, the wage paid by firm A is the sum of the current expected output of the worker at the second-best firm f_{0t} , $y_{f_{0t}}(s_t, k_{0t}) + \varepsilon_{0t}$, and the compensating wage differential $\Psi(s_t, k_{At})$. This term compensates the worker for the three dimensions along which the job offered by firm A differs from the one offered by firm f_{0t} . Specifically, jobs at firms A and f_{0t} may provide different amounts of information about the worker's ability, offer different possibilities of human capital acquisition, and imply different probabilities of exogenous separation. All these dimensions affect the present value of wages from period $t + 1$ on and, due to competition, are reflected in the current wage. This compensating differential would be zero if firms A and f_{0t} provided the worker with identical prospects for information, human capital acquisition, and exogenous separation.

Recall that \bar{U} is the worker's outside option. In equilibrium, when employed by firm A , the worker receives in present value terms the share $\Lambda = (V^w - \bar{U})/(V^A - \bar{U})$ of the match surplus $V^A - \bar{U}$. Note that $\Lambda \in [0, 1]$ since $\bar{U} \leq V^w \leq V^A$. From (13) and the definition of V^w it follows

$$V^w(s_t, \varepsilon_t) = V^0(s_t, \varepsilon_t) - \delta \eta_{Ak_{At}}(h_t) \int_{\varepsilon_{t+1}} E\Pi^0(s_{t+1}, \varepsilon_{t+1}|s_t, k_{At}) dG$$

by simple algebra. Thus, as firm f_{0t} 's match surplus value V^0 varies relative to firm A 's, so does the worker's share of firm A 's match surplus. In particular, the worker extracts the entire surplus of the match with firm A when the best outside offer is from a firm with the same technology as A . In this case, in fact, $\Pi^0 = \Pi^A = 0$ and $V^w = V^0 = V^A$.

The expression in (13) implies a flexible relation between a worker's wage and expected output at firm A . Suppose, for example, that expected output at firm f_{0t} 's job is smaller than at firm A 's job but that firm f_{0t} 's job implies learning, human capital, and separation prospects similar to firm A 's job. Then, the compensating wage differential is (close to) zero and the worker is paid only a fraction of his expected output at firm A : $w_A(s_t, \varepsilon_t) < y_A(s_t, k) + \varepsilon_{Akt}$. By (13), the wage can also more than exhaust the worker's expected output at firm A , that is, $w_A(s_t, \varepsilon_t) \geq y_A(s_t, k) + \varepsilon_{Akt}$, by logic similar to the one in the example in Section 1.1 or when Ψ is large.

3 Data

The data, collected and first analyzed by BGH, are personnel records of all management employees of a medium-sized U.S. firm in a service industry, as of December 31 of each year between 1969 and 1988. I use information on a manager’s year of entry, age, education, job level, (real) salary, and job performance rating.

BGH constructed the firm’s hierarchy using managers’ job titles and transitions across them. It consists of two distinct parts, Levels 1–4, which employ 97.6 percent of managers, and Levels 5–8. Level 8 is the highest title of Chairman-CEO. The average age of managers in the data is 39 years with a standard deviation of approximately 10 years. 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, despite the increase in lower-level entry between 1976 and 1985.¹⁴

Given the small number of individuals observed at Levels 3 and higher, in estimation I treat them as a single level. Level 1, Level 2, and this composite Level 3 correspond to jobs A1, A2, and A3 in the model. Performance ratings range from 1, the highest, to 5, the lowest. Ratings of 3 through 5 are a small fraction of all ratings, so 2–5 are merged into a single measure denoted by ‘0’, resulting in a binary classification of a high or low performance, as in the model.

For 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 and who, during the first ten years of tenure in the firm, experienced no change in the recorded number of years of education and have no level information missing. Due to the high separation rate in each year and tenure (see Table 1), I restrict attention to the first eight years of observations on these managers. In the estimation sample, no individual is older than 45 years at entry. (I also estimated the model on a larger sample that includes entrants into the firm at the original Levels 2, 3, and 4. This larger sample displays very similar features to those of the sample of entrants at Level 1. See the Supplementary Appendix.)

Table 1 displays two patterns about level assignments at the firm and separation from the firm in the estimation sample. First, the percentage of managers at Level 1 rapidly decreases with tenure whereas the percentage of managers assigned to Levels 2 and 3 first increases then decreases with tenure. Second, separation rates are high in all tenures. By the seventh year, over half of those hired at Level 1 have separated. Table 2 converts these distributions into level-specific hazard rates of separation, retention at the same level, and promotion to the next level by tenure. Two further features emerge. First, the separation hazards are approximately constant in tenure. Second, the

¹⁴Baker, Gibbs, and Holmström (1994b) report that the proportion of minorities and women at the firm has increased steadily. My copy of the data does not include information on gender or race.

promotion hazards initially increase then decrease. All promotions are by one level and no demotion is ever observed.

These features of the sample are consistent with a gradual learning process by which managers are progressively assigned to higher levels or separate. That no manager is ever demoted even after repeated low performance provides support for the presence of human capital accumulation. The intuition is that acquired human capital offsets the adverse impact of low performance on beliefs that a manager is of high ability. So, managers are retained at a level even after repeated low performance. Further, allowing human capital to be transferable across levels can also produce flexible tenure profiles for the hazard rate of promotion. Depending on how valuable human capital acquired at a level or overall with tenure is at higher levels, promotion rates may increase or decrease or be non-monotone with tenure in a level or with tenure in the firm. For instance, suppose that tenure in a given level or in the firm makes a manager more productive at a level but not at any other level. This task-specific property of acquired human capital would imply that the probability of promotion out of that level decreases with tenure, regardless of the number of past high ratings, as the greater task-specific human capital a manager acquires makes the manager better suited to that level. As argued in Section 5, the tenure profile of the promotion hazard is crucial for identifying the transferability of acquired human capital across levels.

Table 3 shows two patterns about performance ratings. (I omit statistics on ratings at Level 3 since I do not estimate the parameters of error in performance ratings at this level. See the Appendix for details.) First, the percentage of high ratings at Levels 1 and 2 decreases with tenure. Second, in any given tenure, high ratings are more likely at Level 2 than at Level 1. As shown in the Supplementary Appendix in Table A.15, the percentage of high ratings are also higher among promoted managers than among unpromoted managers and, for promoted managers, among managers promoted early in their tenure in a level relative to those promoted later. These features reinforce the notion that managers are promoted based on their recorded performance. Since in the model human capital is acquired regardless of performance, the greater the sensitivity of the promotion hazard to past high performance, the greater the importance of learning. See Section 5.

Table 4 displays the distribution of wages (in 1988 U.S. dollars) by level and tenure, and Table 5 reports statistics on the distribution of wage changes by tenure. Four features emerge from Tables 4 and 5. Wages are higher at higher levels (note that the percentage of managers earning wages above \$40K increases with the level) and the level distribution of wages is more spread out at higher levels. Managers with higher performance ratings also receive higher wages on average; see BGH. Negative wage changes are frequent: over 20 percent at each tenure. Although average wages increase with tenure, their growth rate is not monotone.

To summarize, average wages increase with tenure and high performance, promotions sometimes occur after low performance, no demotions occur but wage decreases are common, especially after low performance. To account for these features, the model relies on the combination of the human

capital and learning components as neither alone can reproduce these patterns. The human capital acquisition component of the model can naturally lead to increases in average wages with tenure that are unrelated to performance. Moreover, it can explain why promotions sometimes occur after low performance and why repeated poor performance does not lead to demotion. The human capital component, however, cannot generate wage decreases upon low performance. The learning component of the model, instead, naturally gives rise to wage increases upon high performance and wage decreases upon low performance. The learning component, however, also leads to demotions upon low performance, which are not observed in the data.

4 Empirical Specification

Here, I describe how I introduce unobserved heterogeneity into the model and the empirical specification of the output, human capital, and wage processes. In estimation I allow for classification error in performance ratings adapted from Keane and Sauer (2009) and measurement error in wages. Omitted details of the model’s specification are contained in the Appendix and Supplementary Appendix. I also estimated the model on a larger sample that contains observations on managers entering into the firm at levels higher than Level 1. The estimates of the key parameters across the two samples are very similar if not indistinguishable and the model fit is improved. See the Supplementary Appendix for details. Since I observe managers continuously employed at one firm only, from now on the subscript t will denote tenure in the firm.

4.1 Observed and Unobserved Heterogeneity

I assume that the initial prior about ability varies unobservably across managers as do some parameters of output and of the wage distribution. This variation is meant to capture persistent productivity differences among managers, observed by all model agents but not by the econometrician, that may be reflected in observed assignments and paid wages.¹⁵ Formally, in the spirit Heckman and Singer (1984), each manager is of unobserved skill type $i = 1, \dots, I$. In light of changes in likelihood values and the Akaike information criterion, I set $I = 4$. Accordingly, I assume that the initial prior belief that any manager is of high ability takes on value p_{i1} with probability q_i , $i = 1, \dots, I$. Hence, since the probability of entry into the firm is a function of initial prior beliefs, whose distribution I estimate, sample selection is explicitly controlled for. The assumption implicit in this correction, consistent with equilibrium, is that unmeasured determinants of the initial probability of assignment, and thus of entry into the sample, are pure noise conditional on the distribution of the initial prior. These unmeasured determinants reflect idiosyncratic variation

¹⁵Recall that the data do not contain information on managers once they leave the firm. I control for this right censoring by allowing for both endogenous separations (that is, separations dependent on the prior) and exogenous separations (that is, separations independent of ability or performance).

in match quality from period to period, which are captured by the job-specific productivity shocks.

I assume that human capital is h_1 in $t = 1$, captured by a manager's age, education, and year of entry into the firm, and $h_t = h(h_1, t - 1, k_{t-1})$ in $t > 1$, where k_{t-1} denotes a manager's assignment in $t - 1$, so it accrues deterministically with experience, as in a standard learning-by-doing setup.¹⁶ Thus, for a manager of skill type i , the beginning-of-period state in $t \geq 1$ is

$$s_{it} = (p_{it}, h_t, i) = (p_{it}, h_1, t - 1, k_{t-1}, i).$$

4.2 Output and Human Capital

Recall that human capital only affects a manager's high and low performance, $y_{fHk}(\cdot)$ and $y_{fLk}(\cdot)$, at the job k of each firm f . I assume that high and low performance also depend on i and parameterize them, respectively, as $y_{fLk}(h_t, i) = \gamma_{fk} + \gamma_{Li}h_t$ and $y_{fHk}(h_t, i) = y_{fLk}(h_t) + \gamma_{fHi}(t - 1)$. By the same logic as in (1), a manager's expected output at job k of firm f in t is

$$y_f(s_{it}, k) + \varepsilon_{fkt} = \underbrace{\gamma_{fk} + \gamma_{Li}h_t + \beta_{fk}\gamma_{fHi}(t - 1)}_{b_{fk}(h_t, i)} + \underbrace{(\alpha_{fk} - \beta_{fk})\gamma_{fHi}(t - 1)}_{c_{fk}(h_t, i)} p_{it} + \varepsilon_{fkt}. \quad (14)$$

In this expression, as in (1), the human capital process is subsumed by $b_{fk}(h_t, i)$ and $c_{fk}(h_t, i)$, which change deterministically with a manager's experience at a rate that differs by firm, f , and job, k . Due to data limitations, I parameterize the output and human capital process slightly differently at my firm and at its competitors.

Firm A. Since I do not observe output at my firm, I can only identify output differences based on managers' transitions across levels. Hence, I assume that $\gamma_{Ak} = 0$ and, to conserve on parameters, that all dependence of $y_f(s_{it}, k)$ on i is through p_{it} . Since the empirical probabilities of assignment to any level do not vary much with a manager's entry characteristics, h_1 , conditional on tenure, past assignments, and past performance, in practice the relevant state for h_t reduces to $(t - 1, k_{t-1})$. (When estimating the model on the larger sample that also contains observations on entrants at levels higher than Level 1, I allow for h_1 to vary across managers entering at different levels.) Based on these observations, (14) becomes

$$y_A(s_{it}, k) + \varepsilon_{Akt} = \underbrace{\gamma_L h_t + \beta_{Ak}\gamma_{AH}(t - 1)}_{b_{Ak}(t - 1, k_{t-1})} + \underbrace{(\alpha_{Ak} - \beta_{Ak})\gamma_{AH}(t - 1)}_{c_{Ak}(t - 1)} p_{it} + \varepsilon_{Akt}, \quad (15)$$

where $b_{Ak}(t - 1, k_{t-1})$ depends on a manager's current assignment, k , tenure, $t - 1$, and previous period assignment, k_{t-1} , whereas $c_{Ak}(t - 1)$ depends on only a manager's current assignment, k , and tenure, $t - 1$. I estimated $b_{Ak}(t - 1, k_{t-1})$ and $c_{Ak}(t - 1)$ normalizing the variance of ε_{Akt} to

¹⁶Note that h_t can at most depend on tenure at each of the firm's jobs, since managers are observed continually employed at the firm. In practice, $(t - 1, k_{t-1})$ has proved to be a sufficient statistic for it.

$\pi^2/6$. (For simplicity, from now on I will drop the subscript A and let the subscript t denote the dependence on tenure so that $b_{Ak}(t-1, k_{t-1})$ becomes $b_{kt}(k_{t-1})$ and $c_{Ak}(t-1)$ becomes c_{kt} .)

This simple specification of the output and human capital process allows for different degrees of transferability of the human capital acquired at a level of firm A to other levels of the firm. I define the *human capital acquired at Level k* to be *task-general* if $b_{k't}(k) \geq b_{k't}(k')$ for $k' \neq k$, and *task-specific* otherwise. For example, if the human capital acquired at Level 1 or 2 is at least as productive at Level 3 as the human capital acquired at Level 3, then the human capital acquired at Level 1 or 2 is task-general. On the contrary, an extreme case of task-specific human capital is when the human capital acquired at Level 3 only affects output at Level 3.

This specification is also consistent with different degrees of transferability across levels of the human capital acquired with tenure in the firm. I define the *human capital acquired with tenure* in the firm to be *task-general* if the tenure profiles of the parameters $b_{kt}(k_{t-1})$ and c_{kt} at different levels are positively correlated. (See Gibbons and Waldman (2006) for a similar definition.) For example, if the tenure profiles of these parameters at Levels 1 and 2 are highly correlated, then human capital acquired with tenure is task-general across Levels 1 and 2.

Other Firms. Since my data contain information about one firm only, for firm A 's competitors I first assume that $\gamma_{fHi}(t-1)$ is independent of $t-1$ for simplicity. Second, I maintain that $\beta_{fk}\gamma_{fHi}$ and $(\alpha_{fk} - \beta_{fk})\gamma_{fHi}$ are constant across firms, f , and jobs, k . Thus, (14) for a firm f becomes

$$y_f(s_{it}, k) + \varepsilon_{fkt} = \gamma_{fk} + \underbrace{\gamma_{Li}h_t + \beta_{fk}\gamma_{fHi}}_{b(h_t, i)} + \underbrace{(\alpha_{fk} - \beta_{fk})\gamma_{fHi}}_{c(i)}p_{it} + \varepsilon_{fkt}. \quad (16)$$

4.3 Wages

As shown in (13), Bertrand competition implies that a manager's wage reflects the technology of the best competitor of firm A for that manager. Hence, the extent to which paid wages are affected by the prior about a manager's ability and by a manager's acquired human capital at firm A provides information about the extent to which ability and acquired human capital are transferable across my firm and its competitors. By (13), (16), the properties of extreme value distributions, and equilibrium, the wage of a manager of skill type i at firm A in t is

$$w_{Ait} = \underbrace{b(h_t, i) + c(i)p_{it}}_{\varpi_i(h_1) + \omega_1 \cdot (t-1) + c(i)p_{it}} + \ln \underbrace{\sum_{f \neq A, k \in K^f} \exp \{ \gamma_{fk} + \Psi_{fk}(s_{it}, k_{At}) \}}_{\psi_i(h_1, k_{At}) + \psi(k_{At}) \cdot (t-1) + \psi_{1i}(k_{At}) \cdot p_{it}} + \varepsilon_{0t}. \quad (17)$$

See Corollary 1 in the Supplementary Appendix for a proof. (The term Ψ_{fk} is analogous to Ψ in Proposition 2. The notation Ψ_{fk} simply makes explicit the dependence of the compensating wage differential on the job k of the competitor of firm A with technology f . See the Supplementary Appendix.) As (13) and (17) make clear, if firm A has monopsony power in the labor market, then

a manager's current assignment at firm A affects his wage at firm A *only* to the extent that it affects the compensating wage differential $\Psi_{fk}(s_{it}, k_{At})$ between job k of the competitor of firm A with technology f and the assigned job at firm A , k_{At} .

Parameterization of $b(h_t, i)$. I flexibly specify $b(h_t, i)$, as shown beneath the first two terms of (17), as $\varpi_i(h_1) + \omega_1 \cdot (t-1)$ based on three considerations. First, I assume that initial human capital, h_1 , and tenure, $t-1$, affect $b(\cdot)$ in a linearly separable way for simplicity and consistency with the data. Second, in the data wages at any level do not display any significant dependence on k_{t-1} , conditional on manager characteristics and current and past performance, so I ignore it. Third, to conserve on parameters, I capture the dependence on i flexibly through $\varpi_i(h_1)$ but I assume that ω_1 is independent of i .

Parameterization of Compensating Wage Differentials. I specify the third term of (17) containing Ψ_{fk} , as shown beneath it in (17), semiparametrically as a polynomial in p_{it} whose coefficients capture the dependence of Ψ_{fk} on the other elements of s_{it} and on k_{At} .¹⁷ By the logic of the parameterization of $b(\cdot)$, I specify the coefficient on the term of degree zero of this polynomial as $\psi_i(h_1, k_{At}) + \psi(k_{At}) \cdot (t-1)$ and the coefficients on the terms of degree $r \geq 1$ as $\psi_{ri}(k_{At})$. In estimation, the coefficients $\psi_{ri}(k_{At})$ with $r > 1$ have proved unimportant at all trial parameter values, so I ignored them.

Parameterization of Error. To account for unmeasured aspects of firm and manager behavior, as well as recording error, I allow wages to be contaminated by an additive generalized extreme value disturbance, ε_{kit} , where $k = k_{At}$. The dependence of ε_{kit} on i is to allow for the possibility that managers of different skill types have access to different labor market opportunities, reflected in their wages. Since $u_{kit} \equiv \varepsilon_{0t} + \varepsilon_{kit}$ is logistically distributed, I approximate the distribution of u_{kit} by a normal distribution with mean zero and variance σ_{ik}^2 .

Estimated Wage Equation. From (17), by rearranging terms I obtain

$$w_{Ait} = \varpi_i(h_1) + \psi_i(h_1, k_{At}) + [\omega_1 + \psi(k_{At})] \cdot (t-1) + [c(i) + \psi_{1i}(k_{At})]p_{it} + u_{kit}, \quad (18)$$

which leads to a sharing rule of the match surplus between firm A and a manager that varies for the same manager over time and across managers as a function of the prior, p_{it} , human capital, h_t , and the current assignment at firm A , k_{At} . Now, define $\omega_i(h_1, k_{At}) = \varpi_i(h_1) + \psi_i(h_1, k_{At})$. Let $\omega_{1t}(k_{At}) = \omega_1 + \psi(k_{At})$. Since tenure terms at Levels 2 and 3 have proved negligible and difficult to estimate with any precision, I estimate $\omega_{1t}(k_{At})$ only at Level 1 and denote it simply by ω_{1t} .¹⁸ Let also $\omega_{2i}(k_{At}) = c(i) + \psi_{1i}(k_{At})$. Empirically, the dependence of wages on a manager's current and past performance is similar across levels, so I maintain $\omega_{2i}(k_{At}) = \omega_{2i}$ to limit the number of

¹⁷It is easy to see that γ_{fk} and Ψ_{fk} can be expressed in log form so that the third term of (17) can be rewritten as the log sum of polynomials, which is itself a polynomial.

¹⁸Based on descriptive analysis, tenure terms at Levels 2 and 3 have proved unimportant, conditional on managers' characteristics and current and past performance (proxies for the prior). In terms of notation, since $\omega_{1t}(k_{At})$ is assumed nonzero only at Level 1, $\omega_{1t} = \omega_{1t}(k_{At})I(k_{At} = 1)$.

parameters to estimate. Based on the assumptions on the error term, the fact that the empirical wage distribution at each level is close to a lognormal, and that managers are salaried rather than hourly employees, I interpret the model as determining a process for $\ln(w_{Ait}^o)$, where w_{Ait}^o is the observed wage. Then, from (18) with $k = k_{At}$, I estimate

$$\ln(w_{Ait}^o) = \omega_i(h_1, k) + \omega_{1t} \cdot (t - 1) + \omega_{2i}p_{it} + u_{kit}, \quad (19)$$

$$\omega_i(h_1, k) = \varpi_{0ik} + \varpi_{1k}age + \varpi_{2k}age^2 + \varpi_{3k}edu + \sum_{m=1}^9 \varpi_{ym}I(year = m).$$

Here, *age* and *edu* denote a manager's age and number of years of completed education, respectively, at entry into the firm. The coefficients ϖ_{1k} , ϖ_{2k} , ϖ_{3k} , and ϖ_{ym} capture the degree of transferability across firms of the human capital acquired by managers *before* entry into my firm. Specifically, the term $I(year = m)$ is an indicator function that equals one if, and only if, a manager's year of entry into the firm is 197*m*. The coefficients ϖ_{ym} allow for the possibility that business cycle conditions at the time of hiring are permanently reflected in the wages paid by the firm. BGH find such effects important; see also Waldman (2012). The coefficient ω_{2i} captures the degree of transferability of unobserved ability across firm *A* and its competitors. Based on (15) and (19), below I assess the monopsony power of firm *A* by comparing average paid wages to average estimated output (averaged over productivity shocks).

5 Identification

Here I provide an intuitive overview of the identification of the model. I relegate a more formal discussion to the Supplementary Appendix.¹⁹ Specifically, there I show how repeated information on performance ratings allows me to identify the learning parameters $\{\alpha_{Ak}, \beta_{Ak}\}_{k=1}^3$, even in the presence of classification error in performance ratings, and provide conditions for the nonparametric identification of the processes governing the state variables, that is, the prior and acquired human capital, and the discrete choice variable, that is, level assignment, based on Kasahara and Shimotsu (2009) and Hu and Shum (2012).

Learning. To understand the additional identifying power that information on performance ratings provides, consider first the standard argument for the identification of the initial prior and the learning parameters if only repeated information on level assignment were available. (See Crawford and Shum (2005) for a standard reference in the context of demand estimation.) At each level and tenure, the hazard rates of retention at a level, promotion to a higher level, and demotion to a lower level provide information about the initial prior and $\{\alpha_{Ak}, \beta_{Ak}\}_{k=1}^3$. In particular, the extent to which job switching varies with the number of periods managers have been assigned to a given level helps identify $\{\alpha_{Ak}, \beta_{Ak}\}_{k=1}^3$ whereas differences in the assignment probabilities in early

¹⁹I thank Aureo de Paula, Hiro Kasahara, Kyoo il Kim, and Matt Shum for very helpful discussions.

versus late tenures help identify $\{p_{i1}, q_i\}_i^4$.

Differently from common estimation exercises of learning models, my data also include an additional key piece of information that provides a crucial source of identification for $\{\alpha_{Ak}, \beta_{Ak}\}_{k=1}^3$: the performance ratings of all managers in each year of employment, which are the empirical counterparts of a manager's performance in the model. Formally, the observed distribution of performance at each level is a mixture of the distribution of performance of managers of high and low (unobserved) ability. As shown in Section 1.1, in the absence of classification error, the fractions of managers with a high rating at the end of the first and second year of tenure at Level 1 are sufficient to identify α_{A1} and β_{A1} . By the same argument, the fractions of promoted individuals from Level 1 to 2 with one and two consecutive high ratings at Level 2 identify α_{A2} and β_{A2} . A similar argument applies to Level 3. As shown in the Supplementary Appendix, the same logic holds in the presence of classification error in performance ratings.

Learning vs. Human Capital Acquisition. Recall that human capital evolves deterministically with experience, as in a standard learning-by-doing setup, but has no impact on belief updating. Based on this framework, the crucial feature of the data that allows me to distinguish the two processes, when both are present, is the tenure profile of the hazard rate of promotion and its dependence on a manager's performance. To see how, consider an extreme case in which learning is not present in the data but human capital acquisition is. Recall from (15) that human capital is acquired by a manager independently of performance. Thus, when learning is absent, given a manager's history of level assignments, the probability of promotion to a higher level is invariant to the number of realized high ratings. Instead, as learning becomes more important, the hazard rate of promotion becomes more dependent on the number of past high ratings. Hence, for any given history of level assignments, the sensitivity of the promotion rate to the number of realized high ratings distinguishes learning from human capital acquisition when both processes are present. See the Appendix for a discussion of how the degree of transferability of acquired human capital across levels of my firm and of unobserved ability and acquired human capital across my firm and other firms can be detected, respectively, based on the tenure profile of the hazard of promotion and the correlation between wages and performance.

Wages. Recall from (19) that, conditional on beliefs, (log) wages are determined by a linear semiparametric regression model with a random intercept, $\omega_i(h_1, k) + \omega_{1t} \cdot (t-1)$, which is individual- and level-specific and varies with tenure, and with a random slope, ω_{2i} , on the type-specific prior p_{it} . A large literature examines the nonparametric identification of the distribution of random coefficients in the linear regression model; see Hoderlein, Klemelä, and Mammen (2010) for a recent reference. By a standard moment argument, it is easy to show that the intercept function and the coefficient on the prior can be recovered. Observe that information on wages also helps pin down the distribution of the initial prior. In particular, the skewness of the distribution of (log) wages at Level 1 in the first year of tenure, the number and locations of its modes, and the dispersion of

wage observations around them, provide a further source of identification for $\{p_{i1}, q_{i1}\}$.

6 Estimation Results

Tables 6A through 6I report the values of the parameters, estimated by full-information, full-solution, non-parametric maximum likelihood on the sample of managers who entered into the firm at Level 1—the subscript/superscript A is omitted whenever unambiguous. In estimation the discount factor, δ , is fixed at 0.95. All parameters are significant at the one percent level. I relegate a discussion of model fit and of goodness-of-fit test results to the Appendix. Overall, as Tables 1–4 show, the estimated model successfully captures the patterns of managers’ transitions across the levels of the firm’s hierarchy, which are nonlinear and nonmonotone in tenure, and of separation from the firm as well as the distribution of wages and performance at the main levels at each tenure. (The Supplementary Appendix derives the likelihood function, explains the numerical solution of the model, reports estimates on the sample that includes managers who entered into the firm at levels higher than Level 1, and presents a Monte Carlo exercise that demonstrates parameter identification. There I also explain the inferences that, based on the estimates of the wage parameters, can be drawn about the technology of the competitors of my firm. I find the informativeness of their jobs similar to that of the jobs of my firm.)

Uncertainty and Learning. Three findings emerge from the estimates of the parameters in Tables 6A and 6B. First, the distribution of initial priors, $\{p_{i1}, q_i\}_{i=1}^4$, in Table 6A implies a large degree of uncertainty about a manager’s ability and dispersion in information at hiring. Indeed, the average initial prior probability that a newly hired manager is of high ability is 0.473 but initial priors range from 0.338 for skill type 1 to 0.607 for skill type 4 with a standard deviation of 0.102.

Second, the learning parameters $\{\alpha_{Ak}, \beta_{Ak}\}_{k=1}^3$ in Table 6B (firm A subscript is omitted for simplicity) imply that Level 1 is the most informative, and hence plays an important screening role by allowing the firm to gather information about managers’ ability. By the *Blackwell criterion*, the ordering of jobs by their informativeness, from lowest to highest, is Levels 2, 3, and 1.²⁰ Interestingly, their ordering by the probability of success of a manager of high or low ability, from lowest to highest, is the opposite of their ordering by informativeness: Levels 1, 3, and 2. As in a typical multi-armed bandit problem, then, the firm faces a trade-off between productivity and learning through job assignment.

Third, learning is overall gradual but the speed of learning varies across job levels and skill types. To measure the implied speed of learning at each job, I compute how many years of consecutive high performance it takes to infer with a 90 percent chance that a manager is of high ability. At

²⁰By the Blackwell criterion of informativeness (Blackwell (1951)), job k is more informative than job k' if the posterior beliefs reached after performance is observed at job k at the end of a period second-order stochastically dominate the posterior beliefs reached after performance is observed at job k' .

Level 1 at the average prior it takes 20 years. Starting with the same prior at Level 2 or 3, this process takes 23 years. For managers of the fourth skill type, however, this level of certainty takes only 15, 18, and 17 years at Levels 1, 2, and 3, respectively.²¹ This finding is consistent with the fact that uncertainty at entry into the firm is estimated to be large even for managers who arguably have several years of labor market experience.

Output and Human Capital. From the estimates of the output and human capital parameters in Table 6D, it first emerges that human capital acquired at Levels 2 and 3 is highly task-specific. To see this, note that the difference $b_{1t}(L2) - b_{1t}(L1)$ is large and *negative* in $4 \leq t \leq 7$: managers assigned to Level 2 in $t - 1$ and to Level 1 in t produce $b_{1t}(L2) - b_{1t}(L1)$ *less* in t at any prior than managers assigned to Level 1 in $t - 1$ and t . Thus, demoting a manager from Level 2 to 1 is costly for the firm in medium and high tenures. Similarly, the difference $b_{3t}(L3) - b_{3t}(L2)$ is large and *positive* in $4 \leq t \leq 7$: managers assigned to Level 3 in $t - 1$ and t produce $b_{3t}(L3) - b_{3t}(L2)$ *more* in t at any prior than managers assigned to Level 2 in $t - 1$ and to Level 3 in t . Thus, tenure in Level 3 is valuable for output at Level 3. Second, it appears that human capital acquired with tenure in the firm is highly transferable across levels, at least for the part that affects c_{kt} . To see this, note that the parameters c_{kt} at each Level k display the same (approximately hump-shaped) profile with tenure. For example, c_{2t} starts at zero, peaks at $t = 2$, and essentially declines thereafter. In practice, the tenure profiles of the parameters c_{kt} help the model account for the observed hump-shaped pattern of the hazard rates of promotion out of Levels 1 and 2.

Within-Firm Job Assignment Policy. At the estimated values of the parameters of learning, output, and human capital, the firm's optimal assignment policy implies that retained managers have higher priors than newly hired managers and belief variance is higher at lower levels. Manager selection occurs not only through retention but also through promotion: on average managers at higher levels have higher unobserved ability and acquired human capital.

Wages. The estimates of the wage process in Tables 6F–6I are similar to those in the literature concerning the return to entry age, the size of wage increases at promotion, the convexity of wages in job levels, and the magnitude of wage growth on the job. They provide no evidence that this single firm is anomalous within the broader labor market. For instance, Table 6F implies that promotions lead to sizeable permanent increases in wages: a promotion from Level 1 to 2 implies an increase in annual wages of \$781, whereas a promotion from Level 2 to 3 implies an increase in annual wages of \$5,723.²² So, wages are quite convex in job levels. These results are, qualitatively and quantitatively, in line with much of the literature on internal labor markets. See Gibbons and Waldman (1999a,b), Belzil and Bognanno (2008), and Waldman (2012).

²¹These results are consistent with the findings of Nagypál (2007) on the importance of learning for labor market turnover. From Figure 7 in her paper (where all plots are truncated at the tenth year of tenure), the convergence of beliefs, reflected in the tenure profile of match quality and output, seems to occur past the tenth tenure year.

²²These wage increases are calculated by first averaging across manager skill types the level-specific intercepts of mean log wages and then computing the differences in these values after converting them back from logs to levels.

As for managers' observed characteristics, note that the coefficients on entry age and education in Table 6G imply that human capital acquired before entry (captured by age and education at entry into the firm) is transferable across other firms and my firm. For example, 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 $\varpi_1 + 2\varpi_2(29.71) = 0.0102$), 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 $\varpi_{13} + 2\varpi_{23}(29.71) = 0.0041$) 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).

The estimated tenure term ω_{12} at Level 1 in Table 6G is small (recall that no tenure term is estimated at higher levels), which suggests that human capital acquired at the firm is not very transferable to other firms.²³ Since in the wage equation age and education are fixed at their entry values, this finding implies that wage dynamics is driven almost exclusively by learning and competition.²⁴ I will elaborate on this point below.

As for managers' unobserved characteristics, note that the coefficients $\{\omega_{2i}\}_{i=1}^4$ on the prior in Table 6I are sizeable, positive, and significant, implying that average wages increase with the prior. Thus, unobservable ability is transferable across firms, which suggests that matching based on this ability is positive assortative in the market for managers. Note also that the standard deviation of the wage disturbance, $\{\sigma_{ik}\}_{i=1}^4$, essentially decreases with the level. 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 Levels 1, 2, and 3, pooled across tenures, is \$6,936 at Level 1, \$7,077 at Level 2, and \$8,046 at Level 3. (See the note to Table 4 for the data and Table 7A for the model.) Hence, the model's prediction that wage dispersion grows across levels is generated by the endogenous mechanisms of the model, such as the increased dispersion in information, rather than by idiosyncratic unexplained factors.

Monopsony Power. Overall the estimates confirm the importance of explicitly modeling imperfect competition in the labor market. To see why, note that if my firm behaved as a perfect competitor, then the correlation between average estimated output (averaged over productivity shocks at each job) and average wages would be close to one. Instead, this correlation over the first seven years of tenure in the firm 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; otherwise, the correlation is much lower).

²³Note that $\omega_{1t} = \omega_{12}I(t < 5) + \omega_{15}I(t \geq 5)$ where $\omega_{15} = -\omega_{12}$. See the Appendix.

²⁴The estimated year-of-entry effects in Table 6H are consistent with the recession between 1974 and 1979, which depressed the wages of entrants in those years relative to the wages of those who entered in earlier years.

By (19), this finding implies that wage premia at promotion reflect differences in compensating wage differentials across levels, that is, in the prospect for learning, human capital acquisition, and exogenous separation between my firm and other firms in the market, rather than higher output at higher levels at my firm. To see this, note from Table 6F that the wage intercepts of each manager skill type are higher at higher levels and much higher than the corresponding expected output intercepts, the parameters $b_{kt}(\cdot)$ in Table 6D. Thus, wage premia at promotion capture the fact that managers at higher levels sacrifice relatively more valuable prospects for learning and human capital acquisition at other firms than they do at lower levels.

Wage Growth. Overall the estimates imply an increase in average wages of 19.4 percent over the first seven years of tenure in the firm, corresponding to an average yearly growth rate of 3.2 percent. (Wage growth over the first seven years of tenure is higher when separating managers are excluded, 19.4 percent (balanced panel), than when they are included, 18.5 percent (unbalanced panel), which implies the presence of selection through retention.) This magnitude of 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, 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.²⁵

Lastly, observe that (log) wage growth can be decomposed into the contribution of the three terms on the right side of (19)—the contribution of the fourth term, u_{ikt} , is effectively zero. An interpretation of this decomposition is that the first term, $\omega_i(h_1, k)$, captures the effect of unobserved skills, initial human capital, and job assignment, the second term, $\omega_{1t} \cdot (t - 1)$, captures the effect of human capital acquired with tenure in the firm, and the third term, $\omega_{2i}p_{it}$, captures the effect of learning. Over the first seven years of tenure in the firm, the first term accounts for more than 98 percent of the increase in log wages whereas the second term accounts approximately for -1.6 percent (unbalanced panel) to 0 percent (balanced panel). Based on this decomposition, one might conclude that learning accounts only for a trivial percentage of wage growth on the job.

This interpretation, however, is problematic since job assignment is determined by the processes of learning and human capital acquisition at the firm: the 98 percent contribution of unobserved skills, initial human capital, and job assignment to wage growth includes the dynamics of learning and acquired human capital. These two processes, then, contribute to wage growth over and above what is accounted for by the second and third terms of (19) through their impact on the dynamics of job assignment. In particular, learning has an *indirect* effect on wage growth through its effect on promotions: learning leads higher-ability managers to be more quickly promoted to higher levels,

²⁵Buchinsky et al. (2010) 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 individuals with 5 years of experience and 2 years of tenure, and 0.42 log points in the case of individuals with 15 years of experience and 6 years of tenure. The implied yearly growth rate is between 2.9 percent (from $[\exp(0.42) - 1]100/18$) and 8.7 percent (from $[\exp(0.94) - 1]100/18$).

at which they earn higher wages. I uncover this indirect effect of learning on wages through the counterfactual experiments discussed next.

7 Learning, Human Capital, and Persistent Uncertainty

Here I assess the roles of learning, human capital acquisition, and persistent uncertainty about ability for the observed patterns of jobs and wages at my firm.

The Role of Learning and Human Capital. To evaluate the impact of learning on the dynamics of jobs and wages in the baseline model, I compare the predictions of the baseline model to the case when learning is absent. The *no learning* version of the model assumes that all jobs are uninformative about ability ($\beta_{Ak} = \alpha_{Ak}$ for all k). Table 7A compares statistics for wages in the two models, with and without learning. Overall learning generates faster wage growth and greater wage dispersion with tenure. Over the course of the first seven years at the firm, learning accounts for 26 percent (from $(19.4 - 15.4)100/15.4 = 26.0$ percent) of measured wage growth and 23 percent (from $(8,046 - 6,534)100/6,534 = 23.1$ percent) of the variability of wages at Level 3 (with lower percentages at lower levels), the rest being accounted for by human capital acquisition before entry into the firm and with tenure at the firm.²⁶ A key channel for beliefs to affect wages, as anticipated, is job assignment. Table 7B displays the distribution of managers across levels by tenure in the baseline and no learning models. 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, the fraction of managers at Level 3 is 32 percent in the baseline model and 24 percent in the no learning model. Learning, then, implies more rapid promotions for higher-ability managers and, thus, higher wage growth, given that wages at higher levels are on average higher.

Note also that despite human capital acquired at the firm being essentially firm-specific (recall the estimate of the tenure term of wages), the impact of human capital acquisition on wage growth is substantial and, as is the case for learning, it occurs through the dynamics of job assignment. Human capital acquired with tenure, being transferable across levels, makes managers more productive at higher levels, at which they earn higher wages. Further, as discussed above, human capital acquired at a level essentially prevents demotions from occurring and, thus, reduces the wage loss that managers would otherwise experience upon low performance. Both of these channels imply a positive effect of human capital on wages.

Comparison with the Literature on Learning. I find that the *total* effect of learning on wages

²⁶This is the standard way learning is assessed. (See, for instance, Crawford and Shum (2005).) If, when α_{Ak} and β_{Ak} change, output levels are adjusted so as to keep expected output at each level equal to the baseline value at each prior, the experiment can lead to smaller or higher wage growth. Such an exercise, however, ignores the fact that learning is inherently costly since it occurs through experience. Presumably, this feature of learning also explains why performance monitoring is so imperfect, as the estimates imply. This experience component of learning is central to the incentives that the firm and the managers face when they acquire information.

is sizeable despite its *direct* effect discussed above being small. Note that this finding is consistent with the estimates of the importance of learning in the literature. On the one hand, that the direct effect of learning on wages is quantitatively unimportant, once comparative advantage is taken into account, is in line with the results of Gibbons et al. (2005), Lluís (2005), and Hunnes (2012) (at least for the sample of technical white collar workers that Hunnes considers). In addition, by replicating the analysis of Lluís (2005) and Hunnes (2012) in my data, using the same instrumental variable approach as in Gibbons et al. (2005), I find the impact of learning mostly insignificant; details are available upon request. On the other hand, that the indirect effect of learning may be sizeable is consistent, for instance, with the findings of Nagypál (2007) on the importance of learning for labor market turnover and Papageorgiou (2012) on the importance of learning for occupational mobility. For evidence of the presence of learning in the BGH data, based on a perfectly competitive model without job assignment, see Khan and Lange (2011). For related findings on the importance of learning in output markets, see Erdem and Keane (1996), Crawford and Shum (2005), Ching (2010), and the review by Ching, Erdem, and Keane (2011) of the industrial organization literature.

The Role of Learning Through Experimentation. Next, consider the case when experimentation is precluded, that is, when all jobs are equally informative. I focus on one experiment, referred to as *equal informativeness as Level 1*. Jobs at Levels 2 and 3 are made as informative as at Level 1 by setting α_{Ak} and β_{Ak} , $k = 2, 3$, equal to their estimated values at Level 1 while leaving all other parameters unchanged. Results for the analogous experiments of equal informativeness as Levels 2 and 3 are reported in the Supplementary Appendix (Tables A.12-A.14). Table 7C shows that, without experimentation, nearly all of the managers who do not separate from the firm are quickly assigned to Level 3 and retained there. Indeed, apart from the first two years of tenure, the proportion of managers assigned to Level 1 is very small in each tenure. This finding occurs even though learning takes place slowly over time, as discussed. Thus, the speed of promotions per se is not informative about the speed of learning in a job. Rather, the pace of promotions reflects the relative informativeness or, equivalently, the relative speed of learning at different jobs and, thus, the scope for experimentation. This effect proves to be quantitatively important for job mobility at the firm as well as wage growth. As Table 7A shows, rapid promotions are accompanied by rapid wage growth. By the second year of tenure, without experimentation wage growth is 17.6 percent, nearly twice as large as in the baseline model. By the fourth year of tenure, managers experience a higher wage growth than their cumulative wage growth over seven years in the baseline model—after that, wage growth is fairly flat. Note that over the first seven years of tenure, cumulative wage growth would be 20 percent higher (from $20.1 = (23.3 - 19.4)100/19.4$) in the absence of experimentation. Hence, learning without experimentation would lead to higher wage growth than the wage growth generated by learning with experimentation in the baseline model.

The Role of Persistent Uncertainty. I estimate learning to be a slow process. To assess the role of persistent uncertainty, I conduct two experiments. In the *fast learning at Level 1* case,

jobs at Level 1 are made to be nearly perfectly informative about ability, with $\alpha_{A1} = 0.99$ and $\beta_{A1} = 0.01$, whereas the other parameters are fixed at their baseline values.²⁷ See the Appendix for the case of *fast learning at Level 2*. Table 7A shows the implications for wages and Table 7B for job assignments. Over the first seven years of tenure, faster learning at Level 1 leads wages to grow more than 60 percent compared to approximately 20 percent in the baseline model. Wage dispersion at each level also increases: the standard deviation of wages at Level 3 is over five times larger than in the baseline model. Further, promotions to Level 3 occur more rapidly than in the baseline model: by the fourth year of tenure, 30 percent of managers are already assigned to Level 3. Hence, persistent uncertainty about ability substantially compresses wage levels and wage growth with tenure.

8 Conclusion

This paper is the first to develop and estimate a model integrating job assignment, learning about ability, and human capital acquisition to account for rich patterns of individual careers in firms. Using detailed data on individual job assignments, wages, and performance ratings, I can separately identify the process for learning and for human capital acquisition. These data allow me to recover explicit measures of the uncertainty about workers' abilities at the time of hiring, assess the informativeness of each of the firm's jobs, estimate the speed of learning at each job, and measure the relative contribution of learning and human capital acquisition to wage growth on the job.

My results show that the impact of learning on wage growth, through its indirect effect on the dynamics of job assignment, is much larger than the impact usually documented through reduced-form estimation of the wage process alone. These reduced-form estimates capture only a part of the impact of learning on wages, namely its direct and contemporaneous effect, typically found to be small or insignificant. I argue that the key impact of learning on wages is through its indirect effect on the dynamics of job assignment: learning leads to faster promotions, which in turn lead to higher wages. The combination of the direct and indirect effects leads to a large total effect of learning on wages: learning accounts for more than one quarter of cumulative wage growth on the job over the first seven years of tenure, with the rest essentially from human capital acquisition. The different speed of learning at different jobs, that is, the experimentation component of learning, is a primary determinant of the timing of promotions and wage increases, which implies a much more convex wage growth at low tenures than when the speed of learning is identical across jobs. Hence, the view that learning has only a negligible or insignificant impact on wages needs to be re-assessed.

My exercise contributes to a growing empirical literature (see Ferrall (1997) and Ferrall et al.

²⁷If, when α_{Ak} and β_{Ak} change, output levels are adjusted so as to keep expected output at each level equal to the baseline value at each prior, both experiments lead to even higher wage growth.

(2009)), the goal of which is to shed light on the internal workings of firms and to bridge several areas of study. These areas include 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 Oyer and Shaefer (2011)), and, last, human resource management themes on the relationship between productivity and labor practices inside firms (see Bloom and Van Reenen (2011)). As for any structural model, the insights the model has offered should be subject to further verification. Yet overall, my findings attest to the potential of learning and human capital models, augmented along the dimensions considered, to capture salient features of careers.

A Appendix

This appendix contains omitted details of the model, the empirical specification, and the estimation results. See the Supplementary Appendix for further details. It is available online at https://sites.google.com/site/elenapastorino1econ/careers_supptot.pdf.

A.1 General Model: Omitted Details

Output. Realized output in period t is either $y_{fHk}(h_t) + \varepsilon_{fkt}$ or $y_{fLk}(h_t) + \varepsilon_{fkt}$, where ε_{fkt} is a mean-zero type I extreme value productivity shock with c.d.f. G . In the general model, the ability-dependent component of output, $y_{fHk}(h_t)$ or $y_{fLk}(h_t)$, that is, the worker's performance, is modelled as discrete because it is assumed to be measured in the data by a worker's performance rating, which is essentially 'high' or 'low'. I maintain the assumption of only two types of workers, common in the literature, for its convenience and flexibility. It implies that the probability that a worker is of high ability is a sufficient statistic for beliefs and, thus, conveniently reduces the computational burden of maximum likelihood estimation. Together with the assumption of two values for performance, it also parsimoniously allows for greater flexibility in the process for beliefs compared to, say, a normal prior-signal learning case. In the Bernoulli case, unlike in the normal prior-signal case, the variance of posterior beliefs need not decline deterministically in t . Lastly, note that without productivity shocks, employment and job assignment policies, being Markovian, would be deterministic functions of beliefs. Thus, given that worker types and performance ratings are discrete, productivity shocks guarantee a nondegenerate likelihood function.

Best-Response Value Functions. Given the strategies of all firms, at each state the worker's strategy satisfies the Bellman equation

$$V^w(s_t, \varepsilon_t, \cdot) = \max_{\mathbf{d}} \sum_f d_f \left[(1 - \delta)w_{ft} + \delta \eta_{fk_{ft}}(h_t) \int_{\varepsilon_{t+1}} EV^w(s_{t+1}, \varepsilon_{t+1}, \cdot | s_t, k_{ft}) dG \right], \quad (20)$$

where $V^w(s_t, \varepsilon_t, \cdot) = V^w(s_t, \varepsilon_t, \mathbf{w}_t, \mathbf{k}_t)$, $\mathbf{d} = (d_1, \dots, d_N)$ denotes the collection of the worker's possible acceptance ($d_f = 1$) or rejection ($d_f = 0$) decisions with respect to each firm f 's offer of the wage $w_{ft} = w_f(s_t, \varepsilon_t)$ and the job $k_{ft} = k_f(s_t, \varepsilon_t)$, $\sum_f d_f = 1$, and G is the c.d.f. of the future productivity shocks, ε_{t+1} . The expectation EV^w is over the worker's performance realized at the end of the period. Conditional on the worker's assignment to job k of firm f , the beginning-of-period

state next period is $s_{t+1} = (P_{fHk}(p_t), h_{t+1})$ after $y_{fHk}(h_t)$ is realized and $s_{t+1} = (P_{fLk}(p_t), h_{t+1})$ after $y_{fLk}(h_t)$ is realized. Hence, the worker weighs the current wage and the prospects of information acquisition, human capital accumulation, and exogenous separation in deciding which offer to accept. Given the strategies of the other firms and the worker, at each state the strategy of any firm $f \in \{1, \dots, N\}$ solves the Bellman equation

$$\begin{aligned} \Pi^f(s_t, \varepsilon_t) = \max_{w, k} & \left(d_{ft} \left\{ (1 - \delta)[y_f(s_t, k) + \varepsilon_{fkt} - w] + \delta \eta_{fk}(h_t) \int_{\varepsilon_{t+1}} E\Pi^f(s_{t+1}, \varepsilon_{t+1} | s_t, k) dG \right\} \right. \\ & \left. + \sum_{g \neq f} d_{gt} \delta \eta_{gk_{gt}}(h_t) \int_{\varepsilon_{t+1}} E\Pi^f(s_{t+1}, \varepsilon_{t+1} | s_t, k_{gt}) dG \right), \end{aligned} \quad (21)$$

where $d_{ft} = d_f(s_t, \varepsilon_t, \mathbf{w}_t, \mathbf{k}_t)$ and $d_{gt} = d_g(s_t, \varepsilon_t, \mathbf{w}_t, \mathbf{k}_t)$ are the acceptance decisions of the worker with respect to firm f 's and firm g 's offers, $k_{gt} = k_g(s_t, \varepsilon_t)$ is the job offered by firm g , and $E\Pi^f$ is the expectation over the worker's performance realized at the end of the period, conditional on the assigned job: job k if firm f employs the worker or job k_{gt} if firm g employs the worker. From the term in the second line of (21) it is apparent that, as before, in deciding which wage and job offer to make, a firm takes into account the option value of not employing the worker in the current period and attracting him in some future period.

Match Surplus Value Function. To understand how $w_0(s_t, \varepsilon_t)$ and $V^A(\cdot, f_{0t})$ are determined, let f_{1t} denote the firm offering the second-highest value of wages when firm f_{0t} employs the worker. Note that firm f_{1t} may or may not be firm A . It is always firm A if, for instance, the labor market is a duopoly. In equilibrium, $w_0(s_t, \varepsilon_t)$ is the sum of two terms: the worker's one-period expected output at firm $f_t \in \{f_{0t}, f_{1t}\}$, $y_{f_t}(s_t, k_{f_t}) + \varepsilon_{f_t}$, and

$$\Psi(s_t, k_{e_t}) = \frac{\delta}{1 - \delta} \int_{\varepsilon_{t+1}} [\eta_{f_t}(h_t) EV^{f_t}(\cdot | s_t, k_{f_t}) - \eta_{e_t k_{e_t}}(h_t) E\Pi^{f_t}(\cdot | s_t, k_{e_t}) - \eta_{0t}(h_t) EV^w(\cdot | s_t, k_{0t})] dG, \quad (22)$$

which is the compensating wage differential between the offers of firm f_t and the employing firm $e_t = \{A, f_{0t}\}$ at state (s_t, ε_t) . In (22), if $e_t = A$, then $f_t = f_{0t}$ whereas if $e_t = f_{0t}$, then $f_t = f_{1t}$. Formally, the term $\Psi(s_t, k_{e_t})$ compensates the worker for any difference in information, human capital, and probability of exogenous separation between working today at firm f_t and working today at firm e_t . Let $V^w(s_t, \varepsilon_t, f_{0t})$ denote the worker's (expected present discounted) value of wages from accepting the offer of firm f_{0t} . Observe that if $e_t = A$, then $f_t = f_{0t}$ and $(1 - \delta)w_0(s_t, \varepsilon_t)$ equals

$$(1 - \delta)[y_{f_{0t}}(s_t, k_{0t}) + \varepsilon_{0t}] + \delta \int_{\varepsilon_{t+1}} [\eta_{0t}(h_t) E\Pi^0(\cdot | s_t, k_{0t}) - \eta_{Ak_{At}}(h_t) E\Pi^0(\cdot | s_t, k_{Ak_{At}})] dG,$$

which implies that $V^w(s_t, \varepsilon_t, f_{0t})$ can be expressed as

$$\max_{f \neq A} \left\{ \max_{k \in K^f} \left\{ (1 - \delta)[y_f(s_t, k) + \varepsilon_{fkt}] + \delta \eta_{fk}(h_t) \int EV^f(\cdot | s_t, k) dG \right\} - \delta \eta_{Ak_{At}}(h_t) \int E\Pi^f(\cdot | s_t, k_{At}) dG \right\}.$$

Hence, by the properties of extreme value distributions and the definition of V^A ,

$$V^A(s_t, \varepsilon_t, f_{0t}) = \ln \sum_{f \neq A, k \in K^f} \exp \left\{ (1 - \delta)y_f(s_t, k) + \delta \int_{\varepsilon_{t+1}} [\eta_{fk}(h_t) EV^f(\cdot | s_t, k) \right.$$

$$-\eta_{Ak_{At}}(h_t)E\Pi^f(\cdot|s_t, k_{At})] dG\} + \delta\eta_{0t}(h_t) \int_{\varepsilon_{t+1}} E\Pi^A(\cdot|s_t, k_{0t})dG + (1-\delta)\varepsilon_{0t}. \quad (23)$$

See the Supplementary Appendix. Suppose, instead, that $e_t = f_{0t}$, so $f_t = f_{1t}$. Let k_{1t} denote the job offer of firm f_{1t} , ε_{1t} the associated productivity shock, $1 - \eta_{1t}(h_t)$ the corresponding probability of exogenous separation, and V^1 firm f_{1t} 's match surplus value. In this case, $(1 - \delta)w_0(s_t, \varepsilon_t)$ equals

$$(1 - \delta)[y_{f_{1t}}(s_t, k_{1t}) + \varepsilon_{1t}] + \delta \int_{\varepsilon_{t+1}} [\eta_{1t}(h_t)EV^1(\cdot|s_t, k_{1t}) - \eta_{0t}(h_t)EV^1(\cdot|s_t, k_{0t})] dG,$$

which implies that $V^w(s_t, \varepsilon_t, f_{0t})$ can be expressed as

$$\max_{f \neq f_{0t}} \left\{ \max_{k \in K^f} \left\{ (1 - \delta) [y_f(s_t, k) + \varepsilon_{fkt}] + \delta \eta_{fk}(h_t) \int EV^f(\cdot|s_t, k) dG \right\} - \delta \eta_{0t}(h_t) \int E\Pi^f(\cdot|s_t, k_{0t}) dG \right\}.$$

Thus, by the properties of extreme value distributions and the definition of V^A ,

$$V^A(s_t, \varepsilon_t, f_{0t}) = \ln \sum_{f \neq f_{0t}, k \in K^f} \exp \left\{ (1 - \delta) y_f(s_t, k) + \delta \int_{\varepsilon_{t+1}} [\eta_{fk}(h_t) EV^f(\cdot|s_t, k) - \eta_{0t}(h_t) E\Pi^f(\cdot|s_t, k_{0t})] dG \right\} + \delta \eta_{0t}(h_t) \int_{\varepsilon_{t+1}} E\Pi^A(\cdot|s_t, k_{0t}) dG + (1 - \delta) \varepsilon_{1t}. \quad (24)$$

See the Supplementary Appendix. Note that ε_{0t} and ε_{1t} are identically and independently distributed.

Under perfect competition, duopoly, and the oligopoly discussed in the context of the general model, $V^A(s_t, \varepsilon_t, f_{0t})$ simplifies considerably and reduces to the same expression regardless of whether firm A employs the worker or not at state (s_t, ε_t) . Specifically, under perfect competition, it follows that $K^f = K^A$ for all f 's, so $V^A(s_t, \varepsilon_t, f_{0t}) = V^A(s_t, \varepsilon_t, A)$. (Note that under Bertrand competition, the outcome in which firm A makes zero profits in equilibrium at each state and the worker is paid his expected output at firm A already arises when one other firm has the same technology as firm A . In this case, however, unlike in the perfectly competitive case of identical firms, worker turnover across firms with technology A and $f \neq A$ occurs non-randomly as information and human capital are acquired.) Under duopoly between firms A and B , it follows that $K^B \neq K^A$, so $V^A(s_t, \varepsilon_t, f_{0t}) = V^A(s_t, \varepsilon_t, B) \neq V^A(s_t, \varepsilon_t, A)$. Under duopoly, nonetheless, $V^A(s_t, \varepsilon_t) = V^A(s_t, \varepsilon_t, A)$, which is the value function of a standard single-agent dynamic discrete choice problem. (The proof of this result is in Pastorino and Kehoe (2012) for the case without human capital acquisition and without productivity or separation shocks. Adapting their proof to the current case is straightforward.) Under oligopoly, it follows that $K^f \neq K^A$ for $f \neq A$, each technology f is adopted by at least one firm, and

$$\begin{aligned} V^A(s_t, \varepsilon_t, f_{0t}) &= \max_{f \neq A, k \in K^f} \left\{ (1 - \delta) [y_f(s_t, k) + \varepsilon_{fkt}] + \delta \eta_{fk}(h_t) \int_{\varepsilon_{t+1}} EV^A(\cdot|s_t, k) dG \right\} \\ &= \ln \sum_{f \neq A, k \in K^f} \exp \left\{ (1 - \delta) y_f(s_t, k) + \delta \eta_{fk}(h_t) \int_{\varepsilon_{t+1}} EV^A(\cdot|s_t, k) dG \right\} + (1 - \delta) \varepsilon_{0t} \end{aligned} \quad (25)$$

by the properties of type I extreme value distributions, where ε_{0t} is the productivity shock at the job offered by firm f_{0t} . Without loss, the maximization in (25) is intended over the jobs of firms

with *different* technologies. Using the expression of $V^A(s_t, \varepsilon_t, f_{0t})$ in (25),

$$V^A(s_t, \varepsilon_t) = \max \left\{ \max_{k \in K^A} \left\{ (1 - \delta)[y_A(s_t, k) + \varepsilon_{Akt}] + \delta \eta_{Ak}(h_t) \int_{\varepsilon_{t+1}} EV^A(s_{t+1}, \varepsilon_{t+1} | s_t, k) dG \right\}, \right. \\ \left. \max_{f \neq A, k \in K^f} \left\{ (1 - \delta)[y_f(s_t, k) + \varepsilon_{fkt}] + \delta \eta_{fk}(h_t) \int_{\varepsilon_{t+1}} EV^A(s_{t+1}, \varepsilon_{t+1} | s_t, k) dG \right\} \right\}. \quad (26)$$

See Proposition 3 in the Supplementary Appendix.

The perfectly competitive case, the duopoly case, and the oligopoly case discussed are naturally nested by the match surplus value problem $V^A(\cdot) = \max\{V^A(\cdot, A), V^A(\cdot, f_{0t})\}$ in (26). In the perfectly competitive case, $V^A(\cdot, f_{0t}) = V^A(\cdot, A)$. In the duopoly case, $V^A(\cdot, f_{0t}) \neq V^A(\cdot, A)$ and $V^A(\cdot) = V^A(\cdot, A)$. In the oligopoly case, as is apparent from (25), $V^A(\cdot, f_{0t})$ can be equivalently expressed as $v^A(s_t, 0) + (1 - \delta)\varepsilon_{0t}$, where $v^A(s_t, 0)$ is a deterministic continuous function of s_t . For all other market structures, (26) is an approximation, since (23) and (24) entail different expressions for $V^A(s_t, \varepsilon_t, f_{0t})$. The smaller is Π^f for the third-, fourth-, ..., the N -th best firm relative to Π^A , the closer are (23) and (24), and the better is the approximation.

In estimation I compute the probability of employment at firm A and assignment to any of its jobs based on (26). By standard arguments (see Rust (1987, 1994) and the Supplementary Appendix), it follows that $V^A(s_t, \varepsilon_t) = \max_{0 \leq k \leq K^A} \{v^A(s_t, k) + (1 - \delta)\varepsilon_{kt}\}$, where

$$v^A(s_t, k) = (1 - \delta)y_f(s_t, k) + \delta \eta_{fk}(h_t) \int_{\varepsilon_{t+1}} EV^A(s_{t+1}, \varepsilon_{t+1} | s_t, k) dG$$

for $k \geq 1$ and

$$v^A(s_t, 0) = \ln \sum_{f \neq A, k \in K^f} \exp \left\{ (1 - \delta)y_f(s_t, k) + \delta \eta_{fk}(h_t) \int_{\varepsilon_{t+1}} EV^A(\cdot | s_t, k) dG \right\},$$

which I approximate with a flexible polynomial in the prior. See Section A.2.

A.2 Empirical Analysis

Here I provide details about the empirical specification of the model in terms of the parameters for output and human capital, exogenous separations, performance ratings, and wages. I also provide further details about identification, outline the solution of the match surplus value problem of firm A , and present evidence about model fit and specification based on Pearson's χ^2 test.

Model Parameterization. Altogether, the model has 75 estimated parameters: 38 parameters for the distribution of the initial prior, output, human capital, and performance ratings, and 37 parameters for wages.

1. Output and Human Capital. In estimation I set the parameters of output and human capital at the second-best firm to zero. Hence, the parameters $b_{kt}(k_{t-1})$ and c_{kt} are to be interpreted as differences between the values of these parameters at my firm and at the second-best firm. Whenever any of these parameters is zero, such a parameter is to be interpreted to be the same across my firm and the second-best firm. Recall that the data do not contain any direct measures of output and that, in contrast to models of perfect competition, in my model wages are not informative about output at my firm but rather at the second-best firm. Thus, output and human capital parameters at my firm are identified by managers' transitions across the three job levels at my

firm and by separations—in fact, primarily by transitions across the firm’s levels, given that in the data separations from any level are approximately constant in tenure and not very sensitive to performance. See Table 2.

Note that the number of possible transitions across levels proliferates rapidly as tenure increases but there are few observations for many of these transitions. So, it is not possible to estimate the associated parameters with any precision. In order to limit parameter proliferation, I proceed as follows. I parameterize $b_{kt}(k_{t-1})$ at each level as a piecewise function of tenure,

$$b_{kt}(k_{t-1}) = \gamma_k(k_{t-1})tI(t < 4) + \gamma_{kt}(k_{t-1})I(4 \leq t \leq 7). \quad (27)$$

In the case of transitions with either zero or relatively few observations, I did not estimate any parameter of (27) and normalized the corresponding parameters to zero. For transitions with approximately the same number of observations, I maintained that the associated parameters are identical. In particular, since in the data job transitions are by one level, I allowed $b_{kt}(k_{t-1})$ to be nonzero only when k and k_{t-1} are within one level. So, I set $b_{1t}(L3) = b_{3t}(L1) = 0$ in all tenures. Next, note that for any k_{t-1} , the parameters $b_{kt}(k_{t-1})$ are identified by managers’ transitions across the three levels at firm *A*. Hence, a further normalization is needed. For convenience of interpretation, I set $b_{2t}(k_{t-1}) = 0$ in all tenures and consider Level 2 as the reference level. (An alternative could be to impose a form of symmetry across parameters, for instance, by assuming $b_{k't}(k) = b_{kt}(k')$. Such a restriction, however, is highly rejected by the data, since all promotions are by one level and demotions never occur. That is, transitions are strongly one-sided.) These criteria have led to the following parameterization at each level. At Level 1, since all managers are assigned at entry to Level 1, I normalized b_{11} at 1,000. The combination of rapid promotions to Level 2 and high separation rates in each tenure implies that there are few observations at Level 1 in higher tenures. Thus, I set $b_{1t}(L1)$ at zero for $t \geq 2$. Based on the above criteria, I also set $b_{15}(L2) = b_{16}(L2) = b_{17}(L2)$ and estimated $b_{14}(L2)$ and $b_{15}(L2)$ (as differences from the corresponding $b_{1t}(L1)$). At Level 2, I estimated no b_{kt} parameter, as mentioned. At Level 3, I set $b_{35}(L3) = b_{36}(L3)$ and estimated $b_{34}(L3)$, $b_{35}(L3)$, and $b_{37}(L3)$ (as differences from the corresponding $b_{3t}(L2)$).²⁸

Similarly, in terms of the slope parameters c_{kt} of expected output, I did not estimate any parameters at level and tenure combinations with zero or relatively few observations, and normalized them to zero. Also, for parameters at level and tenure combinations with the same number of observations, I maintained they are identical. Formally, at Level 1 I specified

$$c_{1t} = c_{12} \sum_{\tau=2}^3 I(t = \tau) + c_{14} \sum_{\tau=4}^8 I(t = \tau)$$

in light of the different patterns of promotions out of Level 1 from the fourth year of tenure on. Given the small number of observations at Level 1 in high tenures, I only estimated c_{12} and set c_{14} equal to zero. At Level 2, I estimated the parameters c_{22} , c_{23} , c_{25} , and c_{26} , and set $c_{22} = c_{24}$, and $c_{26} = c_{27}$. Given that no manager is assigned to Level 2 at entry and only a small fraction of managers are assigned to Level 2 in the eighth year of tenure, I set both c_{21} and c_{28} at zero.

At Level 3, since no manager is observed at this level until the third year of tenure, I restricted $c_{31} = c_{32} = c_{33}$. Given that the hazard rates of promotion from Level 1 to 2 and from Level 2 to 3 display similar qualitative features from $t \geq 2$ on, I allowed for common components across the

²⁸The reason is that the model can account for transitions from Level 2 to 3 in two ways: either through newly promoted managers from Level 2 to 3 having high output intercept at Level 3 ($b_{3t}(L2)$) or through managers having high output intercept at Level 3 once they have been at Level 3 for at least one period ($b_{3t}(L3)$). Hence, only the difference in these parameters can be precisely pinned down. If, for instance, there was variation in the data in the level from which managers are promoted into Level 3, these parameters could be estimated precisely.

parameters c_{2t} and c_{3t} at $t \geq 2$ to conserve on parameters. In particular, I specified $c_{3t} = c_{3t} + c_{22}$ at $t \geq 4$. The choice of c_{22} as the ‘benchmark’ parameter at Level 2 is due to the high separation rate from Level 2 and the high promotion rate from Level 2 to 3 in each tenure: output parameters of Level 2 in low tenures can be more precisely pinned down. This formulation has led to $c_{35} = c_{36} = c_{22}$. In sum, at Level 3 I just estimated the parameters c_{31} , c_{34} , c_{37} , and c_{38} .

2. Exogenous Separations. I omit the subscript A for simplicity, since the relevant probabilities are estimated only for firm A . I specify the probability of exogenous separation to depend on a manager’s tenure in the firm and current level assignment, so $\eta_{kt} = \eta_k(h_t)$. I allow only for variation in η_{kt} 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 $\eta_{13} = \eta_{14} + \xi_3$, $\eta_{23} = \eta_{22} + \xi_3$, 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 and onward: none, since $\eta_{1t} = \eta_{17}$, $\eta_{2t} = \eta_{27}$, and $\eta_{3t} = \eta_{2t}$, $t \geq 8$. For simplicity, I refer to these parameters as η_{18} , η_{28} , and η_{38} . So, the separation shock parameters I estimate 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.

3. Performance Ratings. Recall that I interpret a rating of 1 at Level k in the data as corresponding to the realization of high performance, $y_{AHk}(h_t, i)$, and a rating of 0 as corresponding to the realization of low performance, $y_{ALk}(h_t, i)$. As do BGH, I assume that, even if a manager is part of a team, through the observation of various aspects of the production process, the evaluator of the manager is able to determine the manager’s contribution to final output. For instance, through conversations with team members and co-workers, the evaluator can infer how well a particular manager performed his job. I also maintain that the evaluator has available sufficient information to separate, on the basis of realized output, $y_{Aj k}(h_t) + \varepsilon_{A k t}$, $j = H, L$, the manager’s performance, $y_{Aj k}(h_t)$, from the idiosyncratic productivity shock, $\varepsilon_{A k t}$. For example, in the context of a bank, the evaluator understands when the value of the portfolio of loans of a particular loan manager has increased in a year simply due to the fact that companies under his responsibility have experienced an unusually positive return in the year, which is unrelated to the manager’s ability.

Based on Keane and Wolpin (1997) and Keane and Sauer (2009), I allow for error in recorded performance ratings as follows. Let $E_0(L_t^o, t) = \Pr(R_t^o = 1 | R_t = 0, L_t^o, t)$ denote the probability of a recorded high rating, $R_t^o = 1$, when a low rating is realized, $R_t = 0$, at level L_t^o in tenure t . Let $E_1(L_t^o, t) = \Pr(R_t^o = 0 | R_t = 1, L_t^o, t)$ denote the probability of a recorded low rating when a high rating is realized at level L_t^o in tenure t . I specify $E_0(L_t^o, t)$ and $E_1(L_t^o, t)$ as

$$E_0(L_t^o, t) = \frac{\exp\{d_0 + d_2(L_t^o)[tI(L_t^o = 1) + (t-1)I(L_t^o = 2)]\}}{1 + \exp\{d_0 + d_2(L_t^o)[tI(L_t^o = 1) + (t-1)I(L_t^o = 2)]\}}, \quad (28)$$

$$E_1(L_t^o, t) = \frac{1}{1 + \exp\{d_0 + d_1 + d_2(L_t^o)[tI(L_t^o = 1) + (t-1)I(L_t^o = 2)]\}}. \quad (29)$$

I let these errors vary both with a manager’s assigned level, since the evaluation of performance is often job-specific, and with a manager’s tenure in the firm, 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. Since no individual is observed at Level 2 at entry, I allow the errors at Level 2 to depend on $t - 1$ rather than t .

Note that the monotonicity condition required for the identification of the classification error rates corresponds to the restriction that the probability of a reported high rating increases with the probability of a true high rating or, equivalently, $d_1 > 0$, which is satisfied by my estimates. See Hausman, Abrevaya, and Scott-Morton (1998).

To limit parameter proliferation, and based on model diagnostic and fit, in (28) and (29) 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. 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, in 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 Level 3 thus proved problematic. As a result, the estimated parameters governing the distribution of performance ratings are $(\{\alpha_{Ak}, \beta_{Ak}\}_{k=1}^3, d_0, d_2(L1), d_2(L2))$. In this specification, the greater is $d_2(L1)$ or $d_2(L2)$, the greater is the persistence in misreporting. Since the probability of a recorded rating is not a linear function of its true probability, this classification error scheme leads to bias (Keane and Sauer (2009)). This feature accommodates the fact that 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, which may lead to systematic misreporting of actual performance. This tendency is well known and has been discussed by Baker, Jensen, and Murphy (1988), Murphy (1992), and Prendergast (1999), among others. As the estimates in Table 6C show, performance is recorded with error that is persistent over time. Note that this error also captures the difficulty of an evaluator in assessing individual performance based on the available information.

4. *Wages.* I set the coefficients ϖ_{1k} , ϖ_{2k} , and ϖ_{3k} , respectively, on *age*, *age*², and *edu*, equal at Levels 1 and 2. I denote their common value by ϖ_1 , ϖ_2 , and ϖ_3 . As for 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 that estimated parameters for the year-of-entry dummies are $\{\varpi_{ym}\}_{m=5}^9$. As for the remaining coefficients, I assume that the coefficient on tenure at Level 1 is $\omega_{1t} = \omega_{12}I(t < 5) + \omega_{15}I(t \geq 5)$ with $\omega_{15} = -\omega_{12}$, to account for the progressively greater proportion of managers at Level 1 who are paid relatively low wages from the fifth tenure year on. Finally, I assume that the standard error of the shock to (log) wages does not vary across skill types at Level 3. So, the estimated wage parameters are $\{\omega_{0i1}, \omega_{0i2}, \omega_{0i3}\}_{i=1}^4$, ϖ_1 , ϖ_2 , ϖ_3 , ϖ_{13} , ϖ_{23} , ϖ_{33} , ω_{12} , $\{\varpi_{ym}\}_{m=5}^9$, $\{\omega_{2i}\}_{i=1}^4$, $\{\sigma_{1k}, \sigma_{2k}, \sigma_{3k}, \sigma_{4k}\}_{k=1}^2$, and σ_3 .

Identification. I discuss here how the degree of transferability of acquired human capital across levels of my firm and the degree of transferability of unobserved and acquired human capital across firms can be detected in my data.

1. *Transferability of Acquired Human Capital Across Levels of My Firm.* A crucial feature of the data that helps me pin down the transferability of human capital acquired at the firm across levels is the variation of the hazard rate of promotion with tenure and performance. To see why, recall (15) and suppose that tenure in the firm or in a given level makes a manager more productive at that level but not at any other level. This task-specific property of acquired human capital would imply that the probability of promotion of a manager out of a level decreases with tenure, regardless of the number of past high ratings, as the greater task-specific human capital a manager acquires makes the manager better suited to the current level. Suppose, instead, that tenure in the firm or in a level substantially improves a manager's expected output at higher levels. In this case, regardless of the number of past high ratings, the probability of promotion out of a level can increase with tenure in the firm or in a level. Moreover, even after several low ratings, managers may not be demoted to a lower level. Hence, the tenure profile of the hazard rate of promotion at each level for different histories of level assignments and performance ratings provides a set of

moments that help identify the degree of transferability of human capital across levels.

2. *Transferability of Unobserved Ability and Acquired Human Capital Across Firms.* The variability of wages with performance is the key piece of information in the data that helps identify the extent to which unobserved ability is transferable across firms. To see why, note first that the fact that managers with higher performance receive higher wages supports the idea that unobserved ability is transferable across firms. Therefore, at each level the strength of the dependence of average wages on past performance helps identify the degree of transferability of unobserved ability across firms. In estimation, this dependence is reflected in the positive estimated coefficient on the prior in the wage equation.

Next, recall that the transferability across firms of the human capital acquired before entry into my firm is formally captured by the dependence of wages on managers' entry characteristics, since $h_1 = (\text{age}, \text{edu}, \text{year})$. The variability of wages with tenure and previous period assignment is, instead, the key piece of information that helps identify the extent to which human capital acquired at my firm is transferable to other firms. For concreteness, suppose that acquired human capital were close to general across firms. Then, for any given number of past high ratings, by (19) average wages would increase with tenure and previous period assignment. Thus, the dependence of average wages on tenure and previous period level assignment, for varying sequences of realized high ratings, provides a set of moments that identify the degree of transferability to other firms of the human capital acquired at my firm. In estimation, only the effect of a manager's tenure, and only at Level 1, has proved somehow important. These results suggest a low degree of transferability to other firms of the human capital acquired at my firm.

Match Surplus Value Problem. Recall the specification of expected output at firm A . Let $k = 0$ denote the event in which a manager is not employed by the firm. By applying the results of Rust (1987, 1994) and assuming that each productivity shock has variance $\pi^2/6$, it follows

$$V^A(p_{it}, t-1, k_{t-1}, \varepsilon_t) = \max_{0 \leq k \leq 3} \{v^A(p_{it}, t-1, k_{t-1}, k) + (1-\delta)\varepsilon_{Akt}\}$$

where $v^A(p_{it}, t-1, k_{t-1}, k)$, for $1 \leq k \leq 3$ and $2 \leq t \leq 7$, is given by

$$\begin{aligned} & (1-\delta)y_A(p_{it}, t-1, k_{t-1}, k) + \delta\eta_{Akt}[\alpha_{Ak}p_{it} + \beta_{Ak}(1-p_{it})] \ln \sum_{0 \leq k' \leq 3} \exp[v^A(P_{AHk}(p_{it}), t, k, k')] \\ & + \delta\eta_{Akt}[1 - \alpha_{Ak}p_{it} - \beta_{Ak}(1-p_{it})] \ln \sum_{0 \leq k' \leq 3} \exp[v^A(P_{ALk}(p_{it}), t, k, k')] \end{aligned} \quad (30)$$

Here $v^A(P_{Aj}(p_{it}), t, k, k')$, $j = H, L$, incorporates the law of motion for p_{it} and h_t conditional on the assignment to Level k in tenure t . Similar expressions can be obtained for $t = 1$ and $t = 8$. See the Supplementary Appendix.

The model implies that the match surplus value of firm A when $k = 0$ is a continuous function of the prior. Given the lack of direct information in my data on other firms, I specify this continuous function as a flexible polynomial in p_{it} . (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.) The terms of degree zero and one of this polynomial are, respectively, intended to be subsumed in $b_{kt}(k_{t-1})$ and c_{kt} , since they cannot be separately identified from $b_{kt}(k_{t-1})$ and c_{kt} . As mentioned, I set them at zero in estimation and interpret $b_{kt}(k_{t-1})$ and c_{kt} as differences with respect to the corresponding parameters of the second-best firm. In estimation, the terms of degree higher than one of this polynomial have proved impossible to estimate with any precision due to the fact that, conditional on tenure, v^A has not exhibited significant curvature in the prior at all

trial sets of parameter values. So, I ignored them. The probabilities of assignment to Level k , $0 \leq k \leq 3$, in each tenure can be easily computed. See the Supplementary Appendix.

Model Fit. In assessing model fit and conducting the counterfactual experiments, I simulated 4,000 prior realizations per manager, drawn from the estimated nonparametric distribution of initial priors. Table 1 shows that the model tracks remarkably well, qualitatively and quantitatively, the profiles of assignment of managers to Levels 1, 2, and 3, which are nonlinear and nonmonotone in tenure, as well as the tenure pattern of manager separation. Table 2 shows that the model accurately reproduces the feature that outflows from Levels 1 and 2 come from an essentially constant hazard rate of separation and a hazard of promotion that is first increasing, then decreasing with tenure. As Table 3 shows, the model also successfully fits the patterns of performance ratings both at Level 1 (with slight overpredictions in the third and fifth years of tenure) and at Level 2 (except for some discrepancies in the fourth year of tenure). Lastly, Table 4 shows that the model reproduces quite well the distribution of wages at each level and tenure, except for slight discrepancies at Level 3 in the highest tenures. The largest such discrepancies are at Level 3 in the sixth and seventh years of tenure, but this difference is partly due to the high rate of attrition in the sample. Indeed, the fit of the model to the wage data based on the larger sample is substantially better. See the Supplementary Appendix.

One criterion to formally evaluate model fit is the Pearson's χ^2 goodness-of-fit test. I perform this test based on the statistic $s \sum_{j=1}^J \{[g(j) - \hat{g}(j)]^2 / \hat{g}(j)\}$, where $g(\cdot)$ indicates the empirical density function of a given endogenous variable, $\hat{g}(\cdot)$ denotes the maximum likelihood estimate of the density function of that variable, s indicates the number of observations, and J 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 in each tenure, the χ^2 goodness-of-fit test does not reject the model at conventional significance levels in any tenure. In terms of the hazard rates of separation, retention at a level, and promotion 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, in this case the outcome of the test is arguably very much influenced by the small number of observations at Levels 1 and 2 in high tenures. In terms of the distribution of performance ratings at Levels 1 and 2 in each tenure, the test does not reject the model at conventional significance levels in any tenure. Finally, as for the distribution of wages at Levels 1, 2, and 3 in 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 in high tenures relative to the size of observations on managers in the first year. See the estimates on the larger sample for improved fit.

The Role of Persistent Uncertainty. In the *fast learning at Level 2* case, jobs at Level 2 are made to be nearly perfectly informative about ability, with $\alpha_{A2} = 0.99$ and $\beta_{A2} = 0.01$, whereas the other parameters are fixed at their baseline values. Similarly, fast learning at Level 2 yields higher wage growth, much larger wage dispersion at Level 2, and faster promotions. Perhaps surprisingly, it also leads to a lower percentage of managers assigned to Level 3 in high tenures and much

lower wage dispersion at Level 3 relative to the baseline model. The reason is that increasing α_{A2} and decreasing β_{A2} makes Level 2 more attractive for managers with high priors, who tend to be managers with high tenures in the firm. In turn, since managers reach Level 3 at lower priors than in the baseline case, the standard deviation of wages at Level 3 is much lower than at Level 2 (and lower than in the baseline model).

References

- Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin** (2011): “Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics”, mimeo.
- 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.
- Chiappori, P.-A., B. Salanié, and J. Valentin** (1999): “Early Starters versus Late Beginners,” *Journal of Political Economy* 107(4), 731–760.
- 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.

- 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.
- Felli, L., and C. Harris** (1996): “Learning, Wage Dynamics, and Firm-Specific Human Capital”, *Journal of Political Economy* 104(4), 838–868.
- Felli, L., and C. Harris** (2004): “Firm-Specific Training”, STICERD Theoretical Economics, Discussion Paper No. TE/04/473.
- Ferrall, C.** (1997): “Empirical Analysis of Occupational Hierarchies”, *Journal of Human Resources* 32(1), 1–34.
- Ferrall, C., K.G. Salvanes, and E.Ø. Sørensen** (2009): “Wages and Seniority When Coworkers Matter: Estimating a Joint Production Economy Using Norwegian Administrative Data”, QED Working Paper 1200.
- Flinn, C.J.** (1986): “Wages and Job Mobility of Young Workers,” *Journal of Political Economy* 94(3), 88–110.
- 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.
- Greenwald, B.** (1986): “Adverse Selection in the Labour Market,” *Review of Economic Studies* 53(3), 325–347.
- Groes, F., P. Kircher, and I. Manovskii** (2010): “The U-Shapes of Occupational Mobility”, mimeo.
- 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.
- Hoderlein, S., J. Klemelä, and E. Mammen** (2010): “Analyzing the Random Coefficient Model Nonparametrically,” *Econometric Theory* 26(3), 804–837.
- 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.
- Hu, Y., and M. Shum** (2012): “Nonparametric Identification of Dynamic Models with Unobserved State Variables,” mimeo, Johns Hopkins University, forthcoming, *Journal of Econometrics*.
- Hunnes, A.** (2012): “Testing the Role of Comparative Advantage and Learning in Wage and Promotion Dynamics”, *International Journal of Manpower* 33(5), 556–582.
- Jovanovic, B.** (1979): “Job Matching and the Theory of Turnover,” *Journal of Political Economy* 87(5), 972–990.

- Jovanovic, B., and Y. Nyarko** (1997): “Stepping-Stone Mobility,” *Carnegie-Rochester Conference Series on Public Policy* 46(1), 289–325.
- Kasahara, H., and K. Shimotsu** (2009): “Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices,” *Econometrica* 77(1), 135–175.
- 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.
- Khan, L. and F. Lange** (2011): “Employer Learning, Productivity and the Earnings Distribution: Evidence from Performance Measures”, mimeo.
- 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. Bruns Eds., Boston: Harvard University Press, 183–215.
- 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.
- Magnac, T., and D. Thesmar** (2002): “Identifying Dynamic Discrete Decision Processes,” *Econometrica* 70(2), 801–816.
- Miller, R.A.** (1984): “Job Matching and Occupational Choice,” *Journal of Political Economy* 92(6), 1086–1120.
- 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.
- Papageorgiou, T.** (2012): “Learning Your Comparative Advantage”, mimeo.
- Pastorino, E.** (2013): “Job Matching Within and Across Firms”, mimeo.
- Pastorino, E., and P.J. Kehoe** (2012): “Efficiency and Pricing in Markets with Experience Goods”, 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.
- Rubinstein, Y., and Y. Weiss** (2007): “Post Schooling Wage Growth: Investment, Search and

- Learning”, in *Handbook of the Economics of Education*, Volume I, E.A. Hanushek and F. Welch Eds., Amsterdam: North-Holland, 1–67.
- 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.
- Sanders, C., and C. Taber** (2012): “Life-Cycle Wage Growth and Heterogeneous Human Capital”, *Annual Review of Economics* 4, 399–425.
- Sauer, R.M.** (1998): “Job Mobility and the Market for Lawyers,” *Journal of Political Economy* 106(1), 147–171.
- Thomas, J.P., and T. Worrall** (1988): “Self-enforcing Wage Contracts,” *Review of Economic Studies* 55(4), 541–54
- Thomas, J.P., and T. Worrall** (2007): “Limited Commitment Models Of The Labour Market,” *Scottish Journal of Political Economy* 54(5), 750–773.
- 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.
- (2012): “Theory and Evidence in Internal Labor Markets,” in *The Handbook of Organizational Economics*, R. Gibbons and J. Roberts, Eds., Princeton, NJ: Princeton University Press.

FIGURE 1. Static Expected Output and Statically Optimal Policies

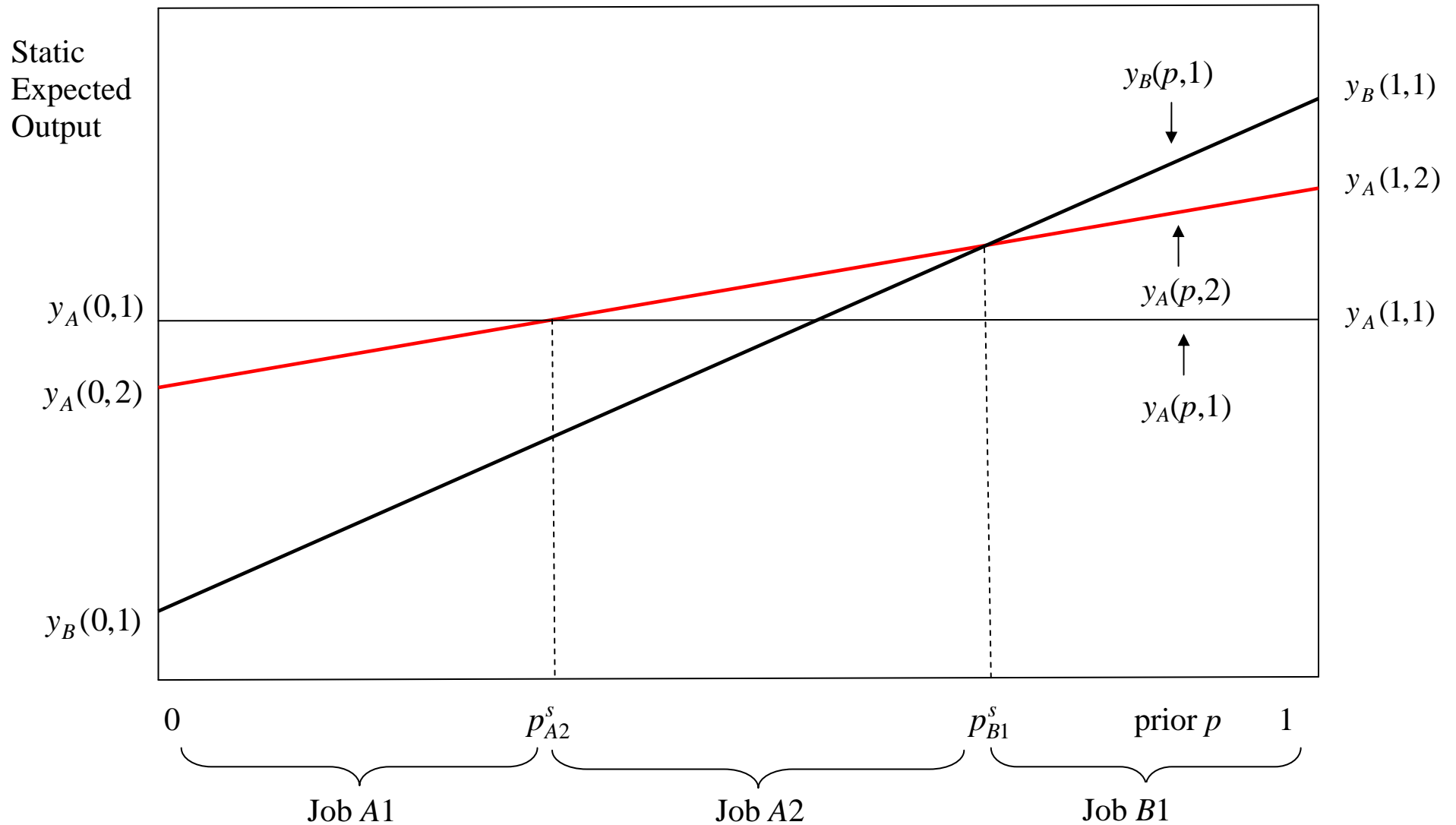
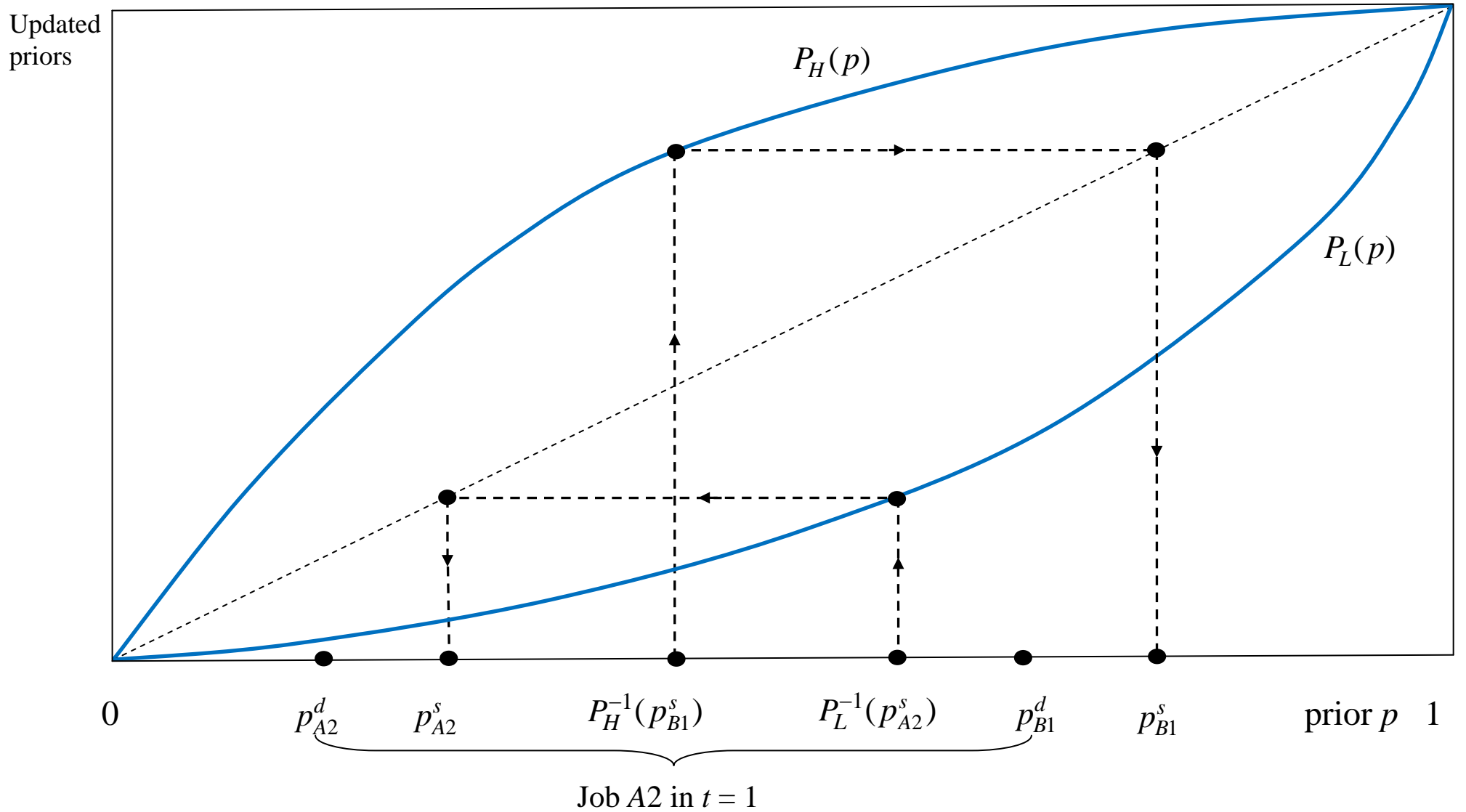


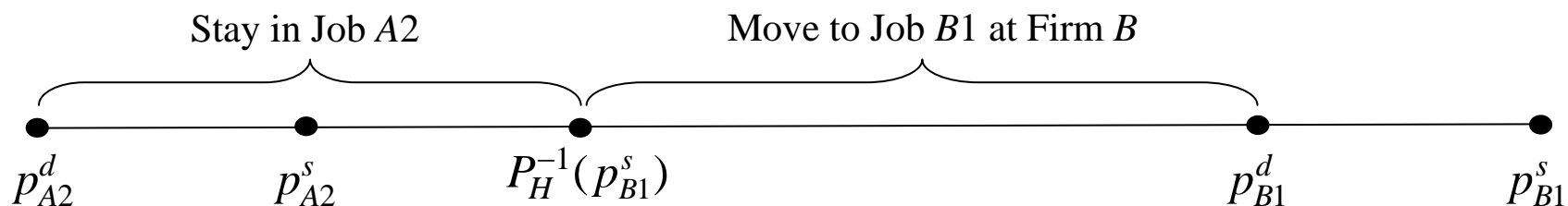
FIGURE 2. Bayesian Updating in Job A2



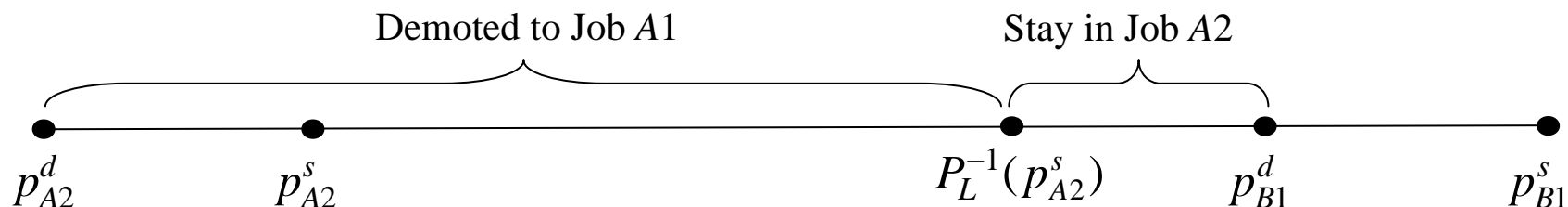
*Note three conditions hold: $p_{A2}^s \leq P_H(p_{A2}^d) < p_{B1}^s$, $P_L(p_{B1}^d) \geq p_{A2}^s$, and $P_H(p_{B1}^d) \geq p_{B1}^s$

FIGURE 3. Jobs Assigned in Period 2 After Job A2 in Period 1*

A. Period 2 Assignment After Success in Job A2 in Period 1



B. Period 2 Assignment After Failure in Job A2 in Period 1



*All workers with priors between p_{A2}^d and p_{B1}^d work in job A2 at firm A in period 1

TABLE 1
Percentage Distribution of Managers Across Levels by Tenure

Tenure	Separation		Level 1		Level 2		Level 3	
	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 2
Hazard Rates of Separation, Retention at Level, and Promotion (Percentages)

Level	Tenure	Separated		Retained		Promoted	
		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 3
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 4
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

*Mean of wages across tenures: \$39,584 at Level 1, \$43,179 at Level 2, \$48,963 at Level 3. Standard deviation of wages across tenures: \$6,924 at Level 1, \$7,377 at Level 2, \$7,270 at Level 3

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 6A
Estimates of Prior Distribution*

Parameters	Type 1 ($i = 1$)	Type 2 ($i = 2$)	Type 3 ($i = 3$)	Type 4 ($i = 4$)
Prior: φ_{i1}	−0.672 (0.022)	−0.484 (0.021)	−0.141 (0.017)	0.435 (0.022)
Mass: q_i	0.155 (0.017)	0.211 (0.030)	0.313 (0.076)	NA

*I let $p_{i1} = \exp\{\varphi_{i1}\} / [1 + \exp\{\varphi_{i1}\}]$ and estimate φ_{i1} : $p_{11} = 0.338$, $p_{21} = 0.381$, $p_{31} = 0.465$, $p_{41} = 0.607$. Asymptotic standard errors in parentheses

TABLE 6B
Estimates of Learning Parameters*

Parameters	Level 1 ($k = 1$)	Level 2 ($k = 2$)	Level 3 ($k = 3$)
High Ability: α_k	0.514 (0.062)	0.5437 (0.006)	0.5435 (0.007)
Low Ability: β_k	0.456 (0.014)	0.491 (0.013)	0.490 (0.010)

*Asymptotic standard errors in parentheses

TABLE 6C
Estimates of Classification Error in Performance Ratings*

Parameters	Level 1 ($k = 1$)	Level 2 ($k = 2$)	Level 3
Base error: d_0	0.521 (0.040)	(same) (same)	NA
Persistence: $d_2(Lk)$	-0.703 (0.040)	-0.544 (0.029)	NA

*See Appendix A.2. Asymptotic standard errors in parentheses

TABLE 6D
Estimates of Output and Human Capital: Intercepts ($b_{kt}(k_{t-1})$) and Slopes (c_{kt})*

Parameters	Value	St. Error
Level 1 ($k = 1$)		
$b_{14}(L2) - b_{14}(L1)$	-704.735	3.577
$b_{15}(L2) - b_{15}(L1)$	-479.607	3.232
c_{12}	2,960.515	13.719
Level 2 ($k = 2$)		
c_{22}	2,563.449	5.570
c_{23}	2,210.102	8.950
c_{25}	2,171.916	8.388
c_{26}	2,224.791	5.423
Level 3 ($k = 3$)		
$b_{34}(L3) - b_{34}(L2)$	853.477	5.941
$b_{35}(L3) - b_{35}(L2)$	202.791	4.475
$b_{37}(L3) - b_{37}(L2)$	228.069	2.715
c_{31}	-399.955	9.659
c_{34}	2,963.404	15.229
c_{37}	2,190.704	1.041
c_{38}	2,003.340	1.066

*See Appendix A.2: $b_{11} = 1,000$, $b_{15}(L2) - b_{15}(L1) = b_{1t}(L2) - b_{1t}(L1)$, $t = 6, 7$; $b_{2t}(k_{t-1}) = 0$, $c_{24} = c_{22}$, $c_{27} = c_{26}$; $b_{35}(L3) - b_{35}(L2) = b_{36}(L3) - b_{36}(L2)$, $c_{33} = c_{32} = c_{31}$, $c_{35} = c_{36} = c_{22}$. Due to the small number of job transitions at certain level/tenure combinations, all other parameters have been set equal to their values at the second-best competitor, which are normalized to zero

TABLE 6E
Estimates of Separation Shocks $(1 - \eta_{kt})^*$

Parameters	Value	St. Error
Level 1 ($k = 1$)		
η_{11}	0.145	0.004
$\eta_{13}(\xi_3)$	0.033	0.001
η_{14}	0.050	0.0001
Level 2 ($k = 2$)		
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
Level 3 ($k = 3$)		
η_{31}	0.122	0.002

*See Appendix A.2: $\eta_{12} = \eta_{11}$, $\eta_{13} = \eta_{14} + \xi_3$, $\eta_{1t} = \eta_{14}$, $t > 4$; $\eta_{22} = \eta_{21}$, $\eta_{23} = \eta_{22} + \xi_3$, $\eta_{2t} = \eta_{27}$, $t > 7$; $\eta_{3t} = \eta_{31}$, $t = 2, 3$, $\eta_{3t} = \eta_{2t}$, $t > 3$

TABLE 6F
Estimates of Wages: Type-Specific Intercept (ϖ_{0ik}) of the Wage Equation*

Parameters	Type 1 ($i = 1$)	Type 2 ($i = 2$)	Type 3 ($i = 3$)	Type 4 ($i = 4$)
Level 1: ϖ_{0i1}	8.805 (0.005)	9.288 (0.005)	9.213 (0.011)	8.865 (0.013)
Level 2: ϖ_{0i2}	8.969 (0.004)	9.359 (0.004)	9.281 (0.009)	8.945 (0.012)
Level 3: ϖ_{0i3}	9.534 (0.008)	9.813 (0.004)	9.738 (0.007)	9.418 (0.011)

*See Appendix A.2. Asymptotic standard errors in parentheses

TABLE 6G
Estimates of Wages: Effect of Age, Education, and Tenure
(Common Across Types)*

Parameters	Value	St. Error
Levels 1 and 2		
Age: ϖ_1	0.028	0.0001
Age ² : ϖ_2	-0.0003	0.000002
Edu: ϖ_3	0.022	0.0004
Level 1		
Tenure: ω_{12}	0.007	0.0003
Level 3		
Age: ϖ_{13}	0.010	0.001
Age ² : ϖ_{23}	-0.0001	0.00001
Edu: ϖ_{33}	0.021	0.001

*See Appendix A.2

TABLE 6H					
Estimates of Wages: Year Dummies ϖ_{ym} (Baseline: 1970-1973)*					
Parameters	1975	1976	1977	1978	1979
	(ϖ_{y5})	(ϖ_{y6})	(ϖ_{y7})	(ϖ_{y8})	(ϖ_{y9})
ϖ_{ym}	-0.063	-0.107	-0.140	-0.208	-0.169
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)

*See Appendix A.2: $\varpi_{y4} = \varpi_{y5}$. Asymptotic standard errors in parentheses

TABLE 6I				
Estimates of Wages: Coefficients on Prior (ω_2) and Standard Deviations (σ_{ik})*				
Parameters	Type 1 ($i = 1$)	Type 2 ($i = 2$)	Type 3 ($i = 3$)	Type 4 ($i = 4$)
Prior: ω_{2i}	2.371 (0.045)	1.833 (0.027)	1.316 (0.015)	1.364 (0.010)
Level 1: σ_{i1}	0.076 (0.001)	0.070 (0.001)	0.057 (0.001)	0.044 (0.001)
Level 2: σ_{i2}	0.063 (0.001)	0.047 (0.001)	0.0302 (0.0004)	0.0303 (0.0004)
Level 3: σ_{i3}	0.047 (0.0004)	(as Type 1) (as Type 1)	(as Type 1) (as Type 1)	(as Type 1) (as Type 1)

*Asymptotic standard errors in parentheses

TABLE 7A
Counterfactual Experiments: Importance of Learning for Wages
Baseline, No Learning, Fast Learning at Level 1, Fast Learning at Level 2*

Statistic	Wages in Each Case				
	Baseline	No Learning	No Experimenta- tion	Fast Learning at Level 1	Fast Learning at Level 2
Means by Level					
Level 1	\$39,584	\$39,706	\$39,763	\$58,271	\$37,847
Level 2	43,179	43,070	42,600	61,451	77,503
Level 3	48,963	48,454	48,818	44,623	24,360
Standard Deviations by Level					
Level 1	\$6,936	\$6,791	\$6,902	\$35,961	\$8,668
Level 2	7,077	6,464	6,831	51,466	45,057
Level 3	8,046	6,534	7,971	45,784	4,281
Cumulative Growth Rates					
Tenure 2	4.6%	3.3%	0.9%	39.3%	8.8%
Tenure 3	8.9	6.8	17.6	48.5	51.5
Tenure 4	13.8	9.8	20.5	52.8	50.1
Tenure 5	15.9	11.1	21.6	55.4	50.3
Tenure 6	17.5	12.9	22.1	58.1	50.1
Tenure 7	18.5	14.6	22.2	60.6	49.8
Tenure 7 (Balanced Panel)	19.4	15.4	23.3	62.5	51.2

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

TABLE 7B
Counterfactual Experiments: Importance of Learning for Level Assignments
Baseline, No Learning, Fast Learning at Level 1, Fast Learning at Level 2*

Tenure	Separation				Level 1				Level 2				Level 3			
	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 : $\alpha_k = 0.99$ and $\beta_k = 0.01, k = 1, 2$

TABLE 7C
Counterfactual Experiment: Importance of Experimentation for Level Assignments
Baseline and Equal Informativeness as Level 1*

Tenure	Separation		Level 1		Level 2		Level 3	
	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: $\alpha_k = \hat{\alpha}_1, \beta_k = \hat{\beta}_1, k = 2, 3$