Federal Reserve Bank of Minneapolis
Research Department Staff Report 470

Revised June 2013

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

Elena Pastorino

Federal Reserve Bank of Minneapolis
and University of Minnesota

ABSTRACT

In this appendix I present details of the model and the empirical analysis, and results of counterfactual experiments omitted from the paper. In Section 1 I describe a simple example that illustrates how, even in the absence of human capital acquisition, productivity shocks, or separation shocks, the learning component of the model can naturally generate mobility between jobs within a firm and turnover between firms. I also include the proofs of Propositions 1 and 2 in the paper. In Section 2 I discuss model identification in detail, where, in particular, I prove that information in my data on the performance ratings of managers allows me to identify the learning process separately from the human capital process. In Section 3 I describe the original U.S. firm dataset of Baker, Gibbs, and Holmström (1994a,b), on which my work is based. In Section 4 I provide details about the estimation of the model, including the derivation of the likelihood function, a description of the numerical solution of the model, and a discussion of the results from a Monte Carlo exercise showing the identifiability of the model's parameters in practice. There I also derive bounds on the informativeness of the jobs of the competitors of the firm in my data, based on the estimates of the parameters reported in the paper. Finally, in Section 5 I present estimation results based on a larger sample that includes entrants into the firm at levels higher than Level 1. Results of counterfactual experiments omitted from the paper are contained in Tables A.12–A.14.

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

*I especially thank Aureo De Paula, Hiro Kasahara, Kyoo il Kim, and Matt Shum for helpful discussions about the identification of the model. I am also grateful to John Geweke for his many useful comments. 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.

1 Omitted Model Details

1.1 An Example

I consider a simple example that illustrates how the model produces nontrivial transitions between jobs within a firm as well as turnover between firms. Although this example implies less rich dynamics than the model in the paper, it is sufficient to clarify several key features about equilibrium in my framework. First, it makes clear that, even if the model does not feature any search frictions and all firms have the same beliefs about a worker's ability, the model implies a nondegenerate distribution of workers to jobs and firms (aside from the limiting case in which all uncertainty about ability is resolved). Second, the example makes clear that the model does not imply perfect short-term assortative matching (outside of the limiting case). Third, it makes clear that the model naturally implies 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 this example, as in the paper, I assume that the market consists of one firm of type (that is, technology) A and at least two firms of type B. Firm A has two jobs, simply referred to as $A1$ and $A2$. Each firm of type B has only one job, simply referred to as $B1$.

1.1.1 Simplest Case

I set up the example so that all of the interesting dynamics occurs for workers who are first assigned to job $A2$. To this end, I assume that job $A1$ is uninformative about worker ability (that is, $\alpha_{A1} = \beta_{A1}$), job $A2$ is moderately informative (with $\alpha_{A2} = \alpha$, $\beta_{A2} = \beta$, and $\alpha, \beta \in (0, 1)$), and job $B1$ is perfectly informative (that is, $\alpha_{B1} = 1$ and $\beta_{B1} = 0$). I also assume that the model has only two time periods and features no human capital acquisition, productivity shocks, or separation shocks. Also, all workers are of the same skill type. Hence, I denote the prior belief that a worker is of high ability in the first period simply by p. I assume that $\alpha > \beta$.

In this simple example, the expected output of the worker at firm $f \in \{A, B\}$ in job k is

$$y_f(p, k) = p\overline{y}_f(\alpha, k) + (1 - p)\overline{y}_f(\beta, k) = \overline{y}_f(\beta, k) + \left[\overline{y}_f(\alpha, k) - \overline{y}_f(\beta, k)\right] p = b_{fk} + c_{fk}p, \quad (1)$$

where $\overline{y}_f(\alpha, k) = \alpha_{fk}y_{fHk} + (1 - \alpha_{fk})y_{fLk}$, $\overline{y}_f(\beta, k) = \beta_{fk}y_{fHk} + (1 - \beta_{fk})y_{fLk}$, $b_{fk} = \overline{y}_f(\beta, k)$, and $c_{fk} = \overline{y}_f(\alpha, k) - \overline{y}_f(\beta, k)$. I assume that parameters are such that

$$y_B(0, 1) < y_A(0, 2) < y_A(0, 1) \text{ and } y_A(1, 1) < y_A(1, 2) < y_B(1, 1). \quad (2)$$

Notice that (2) implies a form of complementarity between ability and jobs: 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 $A1$. Figure 1 illustrates the expected output functions in (1) (averaged over productivity shocks).

Trivially, under (2), if the economy starts with each worker's ability known, so that some workers are known to be of low ability ($p = 0$) whereas others are known to be of high ability ($p = 1$), then the model implies that low ability workers work in job $A1$ of firm A and high ability workers work in job $B1$ of one of the two firms B. That is, the model implies a rather degenerate distribution of workers to jobs and firms, with perfect assortative matching between workers and jobs, and no job mobility within or between firms. The whole point of the learning component of my model, however, is that a worker's ability is imperfectly known. Thus, better matching takes place only over time as information about ability is acquired. (And, as the paper documents, the data point to the existence of substantial initial uncertainty about a worker's ability.)

Consider now the more interesting case in which $p \in (0, 1)$. The match surplus value problem of firm A in the first period is

$$V_1^A(p) = \max \Big\{ \max_{k \in \{1,2\}} \left\{ (1-\delta) y_A(p,k) + \delta \left\{ r_{Ak}(p) V_2^A(P_{AHk}(p)) + [1 - r_{Ak}(p)] V_2^A(P_{ALk}(p)) \right\} \right\},$$

$$(1-\delta) y_B(p,1) + \delta \left\{ r_{B1}(p) V_2^A(P_{BH1}(p)) + [1 - r_{B1}(p)] V_2^A(P_{BL1}(p)) \right\} \Big\},$$

where $r_{Ak}(p) = \alpha_{Ak} p + \beta_{Ak}(1-p)$ and $r_{B1}(p) = \alpha_{B1} p + \beta_{B1}(1-p)$. In this value function, the subscript denotes the time period, the superscript denotes the firm.

I solve for equilibrium starting from the last period, here, period 2. In the last period, the job assignment decision of each firm is static. Clearly, from (2) the equilibrium job assignment policy is to assign job $A1$ at low enough priors, job $A2$ at intermediate priors, and job $B1$ at high enough priors. More formally, 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 $y_A(p_{B1}^s, 2) = y_B(p_{B1}^s, 1)$. From (1) and (2), it follows that $p_{A2}^s = (b_{A1} - b_{A2})/(c_{A2} - c_{A1})$ and $p_{B1}^s = (b_{A2} - b_{B1})/(c_{B1} - c_{A2})$. Hence, the match surplus value in period 2 is

$$V_2^A(p) = \begin{cases} y_A(p,1), \text{ if } p < p_{A2}^s \\ y_A(p,2), \text{ if } p \in [p_{A2}^s, p_{B1}^s) \\ y_B(p,1), \text{ if } p \geq p_{B1}^s. \end{cases} \tag{3}$$

The interesting period is period 1. Observe that the only nontrivial updating rules are for job $A2$. I simplify the notation for them from $P_{AH2}(p)$ and $P_{AL2}(p)$ to

$$P_H(p) = \frac{\alpha p}{\alpha p + \beta(1-p)} \text{ and } P_L(p) = \frac{(1-\alpha)p}{(1-\alpha)p + (1-\beta)(1-p)}.$$

The updating rule for job $A1$ is simply $P_{AH1}(p) = P_{AL1}(p) = p$. The updating rules for job $B1$ are $P_{BH1}(p) = 1$ for $p > 0$ and $P_{BL1}(p) = 0$ for $p < 1$. Thus, the probabilities of high output are given by $r_{A1}(p) = \alpha_{A1} = \beta_{A1}$, $r_{A2}(p) = \alpha p + \beta(1-p)$, and $r_{B1}(p) = p$.

Now consider the first period allocation between jobs $A1$ and $A2$. Since job $A2$ has an informational advantage over job $A1$, the cutoff prior p_{A2}^d at which firm A is indifferent between assigning the worker to jobs $A1$ and $A2$ satisfies

$$p_{A2}^d < p_{A2}^s. \tag{4}$$

Likewise, since job $B1$ has an informational advantage over job $A2$, the cutoff prior p^d_{B1} at which firm A is indifferent between assigning the worker to job $A2$ and having him employed at job $B1$ of a firm of type B satisfies

$$p^d_{B1} < p^s_{B1}. \tag{5}$$

So, a worker with initial prior $p < p^d_{A2}$ starts in job $A1$, a worker with initial prior $p \in [p^d_{A2}, p^d_{B1})$ starts in job $A2$, and a worker with initial prior $p \geq p^d_{B1}$ starts in job $B1$.

To say more than this, I need to determine the job a worker is assigned to after success and failure in the first period. For concreteness, I focus on a region of the parameter space where three conditions hold. First, the worker with the lowest initial prior who is assigned to job $A2$, namely, the worker with prior p^d_{A2}, stays in job $A2$ after a success; that is,

$$P_H(p^d_{A2}) < p^s_{B1}. \tag{6}$$

(Note $P_H(p^d_{A2}) \geq p^s_{A2}$ is already implied by (4); otherwise, p^d_{A2} would equal p^s_{A2}.) At p^d_{A2} a worker who fails is demoted to job $A1$, since $P_L(p^d_{A2}) < p^d_{A2}$ if $\alpha > \beta$ and $p^d_{A2} < p^s_{A2}$ by (4). Thus, $P_L(p^d_{A2}) < p^s_{A2}$. Second, the worker with the highest initial prior at job $A2$ is again assigned to job $A2$ after a failure; that is,

$$P_L(p^d_{B1}) \geq p^s_{A2}, \tag{7}$$

which is to be interpreted as $P_L(p^d_{B1} - \varepsilon) \geq p^s_{A2}$ with $\varepsilon > 0$ arbitrarily small. (Note that $P_L(p^d_{B1}) < p^s_{B1}$ follows from (5) since $p^d_{B1} < p^s_{B1}$.) Third, the worker with the lowest initial prior who is assigned to job $B1$, p^d_{B1}, stays in job $B1$ after a success; that is,

$$P_H(p^d_{B1}) \geq p^s_{B1}. \tag{8}$$

Figure 2 illustrates these assumptions graphically.[1]

Next, I calculate the dynamic cutoff priors. Consider calculating p^d_{A2}, the cutoff prior at which firm A is indifferent between assigning the worker to jobs $A1$ and $A2$ in the first period. This cutoff prior solves

$$y_A(p^d_{A2}, 1) = (1-\delta) y_A(p^d_{A2}, 2) + \delta\{r_{A2}(p^d_{A2}) y_A(P_H(p^d_{A2}), 2) + [1 - r_{A2}(p^d_{A2})] y_A(P_L(p^d_{A2}), 1)\}. \tag{9}$$

The left side of (9) is the value of assigning the worker to job $A1$ in period 1 at prior p^d_{A2}. Here I have used the fact that job $A1$ is uninformative about ability, so the prior is not updated to a different value after either a success or a failure, and the worker stays in job $A1$ in the second period. To see this result, note that a worker assigned to job $A1$ in the first period must have an initial prior $p < p^d_{A2}$. Since $P_{AH1}(p) = P_{AL1}(p) = p < p^s_{A2}$ by (4), the worker is assigned to job $A1$ in period 2 as well.

[1] Observe that the following conditions—$b_{B1} < b_{A2} < b_{A1}$, $b_{A1} + c_{A1} < b_{A2} + c_{A2} < b_{B1} + c_{B1}$, $0 < p^d_{A2} < p^d_{B1} < 1$, $P_H(p^d_{A2}) < p^s_{B1}$, $P_L(p^d_{B1}) > p^s_{A2}$, and $P_H(p^d_{B1}) > p^s_{B1}$—are simultaneously satisfied for the following set of parameters: $\alpha = 0.6$, $\beta = 0.45$, $\delta = 0.1$, $b_{A1} = 3$, $b_{A2} = 2$, $b_{B1} = 0$, $c_{A1} = 1$, $c_{A2} = 5$, and $c_{B1} = 7.5$. Alternatively, these restrictions are satisfied, for instance, for $\alpha \in [0.5, 0.95]$, $\beta = 0.02$, $\delta = 0.95$, $b_{A1} = 3$, $b_{A2} = 2.9$, $b_{B1} = -10.164$, $c_{A1} = 1$, $c_{A2} = 1.2$, and $c_{B1} = 14.414$. By reducing β, the same parameter values would work for values of δ higher than 0.95.

The right side of (9) is the value of assigning the worker to job $A2$ in period 1 at such a cutoff prior. Under (6), the worker is assigned to job $A2$ after a success. In contrast, the worker is assigned to job $A1$ after a failure, since $P_L(p^d_{A2}) < p^s_{A2}$, as argued above.

Consider next the calculation of p^d_{B1}, the cutoff prior at which firm A is indifferent between having the worker at jobs $A2$ and $B1$ in the first period. This cutoff prior solves

$$(1-\delta)y_A(p^d_{B1},2) + \delta\{r_{A2}(p^d_{B1})y_B(P_H(p^d_{B1}),1) + [1-r_{A2}(p^d_{B1})]y_A(P_L(p^d_{B1}),2)\}$$

$$= (1-\delta)y_B(p^d_{B1},1) + \delta[p^d_{B1}y_B(1,1) + (1-p^d_{B1})y_A(0,1)]. \tag{10}$$

The left side of (10) is the value of assigning the worker to job $A2$ in period 1. By (8), the worker is assigned to job $B1$ after a success whereas by (7) and $P_L(p^d_{B1}) < p^s_{B1}$ by (5), the worker is assigned to job $A2$ after a failure. Under these assumptions, the job assignment policy in the first period is

$$\begin{cases} \text{Job } A1 \text{ if } p < p^d_{A2} \\ \text{Job } A2 \text{ if } p \in [p^d_{A2}, p^d_{B1}) \\ \text{Job } B1 \text{ if } p \geq p^d_{B1}. \end{cases}$$

I have set up the example so that the interesting dynamics is generated by workers who start in job $A2$ of firm A in the first period. To see this, let P_H^{-1} and P_L^{-1} denote the inverse functions of P_H and P_L, respectively. After a success, these workers move from job $A2$ to

$$\begin{cases} \text{Job } A2 \text{ if } p \in [p^d_{A2}, P_H^{-1}(p^s_{B1})) \\ \text{Job } B1 \text{ if } p \in [P_H^{-1}(p^s_{B1}), p^d_{B1}), \end{cases} \tag{11}$$

where I have used the facts that $P_H(p^d_{A2}) < p^s_{B1}$ by (6), so the interval $[p^d_{A2}, P_H^{-1}(p^s_{B1}))$ is well-defined, and $P_H(p^d_{A2}) \geq p^s_{A2}$, and the fact that $p^s_{B1} \leq P_H(p^d_{B1})$ by (8), so the interval $[P_H^{-1}(p^s_{B1}), p^d_{B1})$ is well-defined. After a failure, these workers move from job $A2$ to

$$\begin{cases} \text{Job } A1 \text{ if } p \in [p^d_{A2}, P_L^{-1}(p^s_{A2})) \\ \text{Job } A2 \text{ if } p \in [P_L^{-1}(p^s_{A2}), p^d_{B1}), \end{cases} \tag{12}$$

where I have used the fact that $P_L(p^d_{A2}) < p^s_{A2}$ as discussed above, so the interval $[p^d_{A2}, P_L^{-1}(p^s_{A2}))$ is well-defined, and the facts that $p^s_{A2} \leq P_L(p^d_{B1})$ by (7), so the interval $[P_L^{-1}(p^s_{A2}), p^d_{B1})$ is well-defined, and $P_L(p^d_{B1}) < p^s_{B1}$. Figure 3 illustrates these outcomes.

1.1.2 A More General Case

In the more general case, I place no restrictions on the distribution of output at different jobs except that I assume that $\alpha_{fk} > \beta_{fk}$ at each job. So job $A1$ has α_{A1} and β_{A1}, job $A2$ has α_{A2} and β_{A2}, and job $B1$ has α_{B1} and β_{B1}. Since several cases are possible, for concreteness only I continue to assume

(2), so that the static cutoffs continue to satisfy

$$(0 <) p^s_{A2} < p^s_{B1} (< 1)$$

and the job assignment policy in the second period is

$$\text{job } A1 \text{ for } p \in [0, p^s_{A2}), \text{ job } A2 \text{ for } p \in [p^s_{A2}, p^s_{B1}), \text{ and job } B1 \text{ for } p \in [p^s_{B1}, 1].$$

Note that the relation between the static cutoffs p^s_{A2} and p^s_{B1} and the dynamic ones p^d_{A2} and p^d_{B1} depends on the relative informativeness of the jobs. If job $A1$ is more informative than job $A2$, then job $A1$ has an informational advantage over job $A1$, so $p^s_{A2} < p^d_{A2}$ (abstracting from the trivial case of equality between the two cutoffs). Thus, at $p \in [p^s_{A2}, p^d_{A2})$, even though job $A2$ statically dominates job $A1$, assigning job $A1$ in period 1 is still optimal for firm A because the informational advantage of job $A1$ over job $A2$ implies that job $A1$ has a higher (dynamic) match surplus value. In contrast, if job $A2$ is more informative than job $A1$, then the opposite relation holds: $p^d_{A2} < p^s_{A2}$ (again abstracting from the trivial case of equality between the two cutoffs). The same analysis applies to comparing job $A2$ to job $B1$: if $A2$ is more informative than $B1$, then $p^s_{B1} < p^d_{B1}$, whereas if $B1$ is more informative than $A2$, then $p^d_{B1} < p^s_{B1}$.

Observe also that for any given interval of priors at which a given job is assigned in period 1, this interval typically splits into subintervals, which determine a worker's assignment after a success or a failure in the job. To be concrete, consider job $A2$, which is assigned at all initial priors $p \in [p^d_{A2}, p^d_{B1})$. To indicate what happens after a success, as before I divide this interval into two subintervals, a left subinterval $[p^d_{A2}, P^{-1}_H(p^s_{B1}))$, well-defined since $P_H(p^d_{A2}) < p^s_{B1}$ by (6), and a right subinterval $[P^{-1}_H(p^s_{B1}), p^d_{B1})$, well-defined since $P_H(p^d_{B1}) \geq p^s_{B1}$ by (8). In the left subinterval, a success in job $A2$ leads the worker to stay in that job, since even at the highest priors in that interval, which are just below $P^{-1}_H(p^s_{B1})$, a success leads to a posterior below p^s_{B1}. In the right subinterval, instead, a success in job $A2$ leads the worker to move to a firm of type B and work in job $B1$, since even at the lowest prior in that interval, $P^{-1}_H(p^s_{B1})$, a success leads to a posterior equal to p^s_{B1}.

Likewise, to indicate what happens after a failure in job $A2$, I divide the interval into two other subintervals: a left subinterval $[p^d_{A2}, P^{-1}_L(p^s_{A2}))$, well-defined since $P_L(p^d_{A2}) < p^s_{A2}$ as discussed, and a right subinterval $[P^{-1}_L(p^s_{A2}), p^d_{B1})$, well-defined since $P_L(p^d_{B1}) \geq p^s_{A2}$ by (7). In the left subinterval, a failure in job $A2$ leads the worker to be demoted to job $A1$, since even at the highest priors in that interval, which are just below $P^{-1}_L(p^s_{A2})$, a success leads to a posterior below p^s_{A2}. In the right subinterval, a failure in job $A2$ leads the worker to stay in that job, since even at the lowest prior in that interval, $P^{-1}_L(p^s_{A2})$, a failure leads to a posterior equal to p^s_{A2}.

So far I have discussed what happens to workers who start in job $A2$. In this more general case, a new possibility arises for workers who start in job $A1$, those with initial priors $p \in [0, p^d_{A2})$. To determine job assignment after a success here, I split this interval into three subintervals: a left subinterval $[0, P^{-1}_H(p^s_{A2}))$, a middle subinterval $[P^{-1}_H(p^s_{A2}), P^{-1}_H(p^s_{B1}))$, and a right subinterval $[P^{-1}_H(p^s_{B1}), p^d_{A2})$. In the left subinterval, a success leads the worker to stay in job $A1$, in the middle subinterval, a success leads the worker to move to job $A2$, and in the right subinterval, a success leads the worker to turn over to job $B1$ of a firm of type B. Of course, for this to happen, job $A1$ needs to be sufficiently informative. Likewise, workers who start in job $B1$ have three possibilities after a failure: those in a left subinterval turn over to job $A1$ of firm A, those in a middle subinterval turn over to job $A2$ of firm A, and those in a right subinterval stay in job $B1$.

Throughout this example, I have assumed firm A has two jobs and firm B has one job. The more general case, with three or more jobs for firm A and three or more jobs for each of the other firms B, C, D, and so on, each of which faces competition from at least another firm with the same technology, yields to many more cases. In this sense, even this very simple model can generate rich patterns of job mobility within and across firms. When I incorporate into this model human capital acquisition, productivity shocks, and separation shocks, as well as multiple initial priors, the model is flexible enough to produce the rich patterns of mobility across jobs, which are nonlinear and nonmonotone in a manager's tenure, as well as the dynamics of wages observed in the data.

1.2 Equilibrium Characterization

I prove here the main characterization results in the paper in Propositions 1 and 2. I also derive the conditions leading to the specializations of these results in the oligopoly version of the model.

1.2.1 Setup and Equilibrium

Suppose that the market is composed of firms that may all be operating exclusive technologies, $\{B, C, \ldots\}$, which may all imply different learning, human capital acquisition, and exogenous separation possibilities. The equilibrium notion I employ is a refinement of Markov perfect equilibrium, which amounts to a strengthening of its perfection requirement. This notion of Markov perfect equilibrium implies that a firm that does not employ the worker at a state is indifferent between employing and not employing the worker *at the offer it would have made if it had employed the worker.* Formally, I require that if in equilibrium firm f does not employ the worker at state (s_t, ε_t), then

$$\sum\nolimits_{g \neq f} d_{gt} \left\{ \delta \eta_{g k_{gt}}(h_t) \int_{\varepsilon_{t+1}} E\Pi^f(s_{t+1}, \varepsilon_{t+1} | s_t, k_{gt}) dG \right\}$$

$$= \max_{w,k} \left\{ (1-\delta)[y_f(s_t, k) + \varepsilon_{fkt} - w] + \delta \eta_{fk}(h_t) \int_{\varepsilon_{t+1}} E\Pi^f(s_{t+1}, \varepsilon_{t+1} | s_t, k) dG \right\}. \tag{13}$$

This equilibrium notion is a natural generalization to an oligopoly game of wage and job competition of the concept of *cautious Markov perfect equilibrium* introduced by Bergemann and Välimäki (1996). Bergemann and Välimäki consider a dynamic oligopoly game of price competition in which firms are uncertain about consumers' taste for the quality of their products and learn about them over time by observing consumers' experiences. They assume, however, that prices are the only strategic dimension of the competition among firms. In my framework firms compete in prices (wages) as well as jobs. Observe that the notion of equilibrium I use, as well as Bergemann and Välimäki's, is similar in spirit to trembling hand perfection. See Pastorino and Kehoe (2012) for a micro-foundation of this notion.

1.2.2 Main Results

Here I derive three results. I first prove Propositions 1 and 2 in the paper. The third result provides conditions under which these two results specialize to the corresponding ones in the simple version of the model. Recall from the paper that I denote by $k_{0t} = k_0(s_t, \varepsilon_t)$ the job offered by the best competitor of firm A at state (s_t, ε_t), firm $f_{0t} = f_0(s_t, \varepsilon_t)$, firm f_{0t}'s value of profits by Π^0 and match surplus value by V^0, and the worker's expected output at this job by $y_{f_{0t}}(s_t, k_{0t}) + \varepsilon_{0t}$ with associated separation shock $1 - \eta_0(h_t)$. I let $k_{0t} = k_{f_{0t}}(s_t, \varepsilon_t) \equiv k_0(s_t, \varepsilon_t)$ by slight abuse of notation. Hence,

$$V^A(s_t, \varepsilon_t, f_{0t}) = (1-\delta)w_0(s_t, \varepsilon_t) + \delta\eta_0(h_t)\int_{\varepsilon_{t+1}} EV^A(s_{t+1}, \varepsilon_{t+1}|s_t, k_{0t})dG.$$

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*

$$\max\left\{\max_{k\in K^A}\left\{(1-\delta)[y_A(s_t,k) + \varepsilon_{Akt}] + \delta\eta_{Ak}(h_t)\int_{\varepsilon_{t+1}} EV^A(s_{t+1}, \varepsilon_{t+1}|s_t, k)dG\right\}, V^A(s_t, \varepsilon_t, f_{0t})\right\}. \tag{14}$$

Proof. Consider equilibrium states at which firm A employs the worker. I first show that when firm A employs the worker, the job offered by firm A solves

$$\max_{k\in K^A}\left\{(1-\delta)[y_A(s_t,k) + \varepsilon_{Akt}] + \delta\eta_{Ak}(h_t)\int_{\varepsilon_{t+1}} EV^A(s_{t+1}, \varepsilon_{t+1}|s_t, k)dG\right\}. \tag{15}$$

To see this, note that if the worker chooses firm A, it must be that

$$\begin{aligned}&(1-\delta)w_A(s_t, \varepsilon_t) + \delta\eta_{Ak_{At}}(h_t)\int_{\varepsilon_{t+1}} EV^w(s_{t+1}, \varepsilon_{t+1}|s_t, k_{At})dG \\ &\geq (1-\delta)w_f(s_t, \varepsilon_t) + \delta\eta_{fk_{ft}}(h_t)\int_{\varepsilon_{t+1}} EV^w(s_{t+1}, \varepsilon_{t+1}|s_t, k_{ft})dG\end{aligned} \tag{16}$$

for any other firm f, where $k_{ft} = k_f(s_t, \varepsilon_t)$ is the job offered by firm f. Note that (16) must hold with equality when firm f is the second-best firm $f_{0t} = f_0(s_t, \varepsilon_t)$. Otherwise, firm A could increase its profits by modifying, even marginally, its offer.

Consider now the employing firm's problem. Firm A realizes that the worker will accept its wage and job offer only if the offer (w_{At}, k_{At}) is at least as attractive as any other firm's offer. Thus, when firm A employs the worker, it solves the problem

$$\max_{w,k}\left\{(1-\delta)[y_A(s_t,k) + \varepsilon_{Akt} - w] + \delta\eta_{Ak}(h_t)\int_{\varepsilon_{t+1}} E\Pi^A(s_{t+1}, \varepsilon_{t+1}|s_t, k)dG\right\} \tag{17}$$

$$\begin{aligned}&\text{s.t. } (1-\delta)w + \delta\eta_{Ak}(h_t)\int_{\varepsilon_{t+1}} EV^w(s_{t+1}, \varepsilon_{t+1}|s_t, k)dG \\ &\geq (1-\delta)w_f(s_t, \varepsilon_t) + \delta\eta_{fk_{ft}}(h_t)\int_{\varepsilon_{t+1}} EV^w(s_{t+1}, \varepsilon_{t+1}|s_t, k_{ft})dG\end{aligned} \tag{18}$$

for all $f \neq A$, with (18) holding as an equality against firm $f = f_{0t}$. I can then use (18), holding as an equality against firm f_{0t}, to substitute out the wage offer of firm A from (17). I can then rewrite (17) as

$$\max_{k \in K^A} \Bigg\{ (1-\delta)[y_A(s_t,k) + \varepsilon_{Akt}] + \delta\eta_{Ak}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k)dG$$

$$+\delta\eta_{Ak}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k)dG$$

$$-(1-\delta)w_0(s_t,\boldsymbol{\varepsilon}_t) - \delta\eta_0(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})dG \Bigg\}.$$

Dropping irrelevant constants independent of firm A's choices, namely the terms dependent on firm f_{0t}'s choices, I obtain that the job offered by firm A when it employs the worker, $k^A(s_t,\boldsymbol{\varepsilon}_t)$, solves

$$\max_{k \in K^A} \left\{ (1-\delta)[y_A(s_t,k) + \varepsilon_{Akt}] + \delta\eta_{Ak}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k)dG \right\}, \tag{19}$$

since, by definition, $V^A {=} \Pi^A + V^w$. Hence, at states at which firm A employs the worker, using (19) and $k_{At} = k_A(s_t,\boldsymbol{\varepsilon}_t)$, the match surplus value of firm A can be expressed as

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

$$= (1-\delta)[y_A(s_t,k_{At}) + \varepsilon_{Ak_{At}t}] + \delta\eta_{Ak_{At}}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})dG.$$

To prove the rest of the claim, compute now the match surplus value of firm A at states at which firm A employs the worker if, instead of accepting firm A's offer, the worker deviates and accepts firm f_{0t}'s offer. Based on equilibrium strategies, the match surplus value of firm A in this case is equal to

$$(1-\delta)w_0(s_t,\boldsymbol{\varepsilon}_t) + \delta\eta_0(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} [EV^w(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t}) + E\Pi^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})]dG$$

$$= (1-\delta)w_0(s_t,\boldsymbol{\varepsilon}_t) + \delta\eta_0(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})dG$$

by definition of V^A. To interpret this expression, note that firm A's current payoff is zero, the worker's current payoff is the wage $w_0(s_t,\boldsymbol{\varepsilon}_t)$, and both firm A and the worker update beliefs and the worker's human capital based on the job the worker is assigned to at firm f_{0t}. Now, at equilibrium states at which firm A employs the worker, firm A must weakly prefer employing to not employing the worker whereas the worker is indifferent between firm A's and firm f_{0t}'s offers. Hence, by summing firm A's and the worker's values on and off the equilibrium path at states at which firm A employs the worker, it follows

$$(1-\delta)[y_A(s_t,k_{At}) + \varepsilon_{Ak_{At}t}] + \delta\eta_{Ak_{At}}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})dG$$

$$\geq (1-\delta)w_0(s_t,\boldsymbol{\varepsilon}_t) + \delta\eta_0(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})dG. \tag{20}$$

Consider now equilibrium states at which firm f_{0t} employs the worker. The match surplus value of firm A in this case is given by

$$(1-\delta)w_0(s_t,\boldsymbol{\varepsilon}_t) + \delta\eta_0(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})dG$$

since firm A's current payoff is zero and the worker's current payoff is the wage $w_0(s_t,\boldsymbol{\varepsilon}_t)$, by a logic similar to the above one. The match surplus value of firm A if the worker, instead of accepting firm f_{0t}'s offer, deviates and accepts firm A's offer is given by

$$(1-\delta)[y_A(s_t,k_{At}) + \varepsilon_{Ak_{At}t}] + \delta\eta_{Ak_{At}}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})dG$$

where the choice of $k_{At} = k_A(s_t,\boldsymbol{\varepsilon}_t)$ satisfies (15) by (13). Now, at equilibrium states at which the worker accepts firm f_{0t}'s offer, the worker must weakly prefer firm f_{0t}'s offer to firm A's offer whereas firm A, since it is one of the non-employing firms, must be indifferent between not employing and employing the worker. Then, by summing firm A's and the worker's values on and off the equilibrium path at states at which firm f_{0t} employs the worker, I obtain

$$(1-\delta)w_0(s_t,\boldsymbol{\varepsilon}_t) + \delta\eta_0(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})dG$$

$$\geq (1-\delta)[y_A(s_t,k_{At}) + \varepsilon_{Ak_{At}t}] + \delta\eta_{Ak_{At}}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})dG. \tag{21}$$

By combining (19), (20) at states at which firm A employs the worker, and (21) at states at which firm f_{0t} employs the worker, it follows that $V^A(s_t,\boldsymbol{\varepsilon}_t)$ equals (14). $\square$

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

$$w_A(s_t,\boldsymbol{\varepsilon}_t) = y_{f_{0t}}(s_t,k_{0t}) + \Psi(s_t,k_{At}) + \varepsilon_{0t}, \tag{22}$$

where

$$\Psi(s_t,k_{At}) = \frac{\delta}{1-\delta}\int_{\boldsymbol{\varepsilon}_{t+1}} \left[\eta_0(h_t)EV^0(\cdot|s_t,k_{0t}) - \eta_{Ak_{At}}(h_t)EV^0(\cdot|s_t,k_{At})\right] dG. \tag{23}$$

Proof. The worker's indifference between the offers of firm A and firm f_{0t} implies that

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

$$= (1-\delta)w_0(s_t,\boldsymbol{\varepsilon}_t) + \delta\eta_0(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})dG. \tag{24}$$

Now, by (13) in equilibrium firm f_{0t} must be indifferent between its payoff on the equilibrium path, obtained by not employing the worker, and off the equilibrium path, obtained by employing the worker.

Equivalently, with $\Pi^0 \equiv \Pi^{f_{0t}}$ it follows that

$$\delta\eta_{Ak_{At}}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^0(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})dG$$

$$= (1-\delta)\left[y_{f_{0t}}(s_t,k_{0t})+\varepsilon_{0t}-w_0(s_t,\boldsymbol{\varepsilon}_t)\right]+\delta\eta_0(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^0(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})dG.$$

By rearranging terms, I obtain

$$w_0(s_t,\boldsymbol{\varepsilon}_t) = y_{f_{0t}}(s_t,k_{0t})+\varepsilon_{0t}+\frac{\delta}{1-\delta}\int_{\boldsymbol{\varepsilon}_{t+1}}\left[\eta_0(h_t)E\Pi^0(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})\right.$$

$$\left.-\eta_{Ak_{At}}(h_t)E\Pi^0(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})\right]dG. \tag{25}$$

Substituting this last expression into (24) yields that

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

$$-\delta\eta_0(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})dG = (1-\delta)[y_{f_{0t}}(s_t,k_{0t})+\varepsilon_{0t}]$$

$$+\delta\int_{\boldsymbol{\varepsilon}_{t+1}}\left[\eta_0(h_t)E\Pi^0(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})-\eta_{Ak_{At}}(h_t)E\Pi^0(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})\right]dG$$

or, equivalently, by collecting terms,

$$w_A(s_t,\boldsymbol{\varepsilon}_t) = y_{f_{0t}}(s_t,k_{0t})+\varepsilon_{0t}+\frac{\delta}{1-\delta}\int_{\boldsymbol{\varepsilon}_{t+1}}\left\{\eta_0(h_t)[E\Pi^0(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})\right.$$

$$\left.+EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})]-\eta_{Ak_{At}}(h_t)\left[E\Pi^0(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})+EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})\right]\right\}dG.$$

Since, by definition, $V^0 = \Pi^0 + V^w$, the desired result follows. □

By (25) it follows that

$$w_0(s_t,\boldsymbol{\varepsilon}_t) = y_{f_{0t}}(s_t,k_{0t})+\varepsilon_{0t}$$

$$+\frac{\delta}{1-\delta}\int_{\boldsymbol{\varepsilon}_{t+1}}\left[\eta_0(h_t)EV^0(\cdot|s_t,k_{0t})-\eta_0(h_t)EV^w(\cdot|s_t,k_{0t})-\eta_{Ak_{At}}(h_t)E\Pi^0(\cdot|s_t,k_{At})\right]dG. \tag{26}$$

Suppose now that firm f_{0t} employs the worker and firm f_{1t} is the best competitor of firm f_{0t} (note that it may or may *not* be A), with offered job $k_{1t} = k_{f_{1t}}(s_t,\boldsymbol{\varepsilon}_t) \equiv k_1(s_t,\boldsymbol{\varepsilon}_t)$, associated productivity shock ε_{1t}, separation shock $1-\eta_1(h_t)$, value of profits Π^1, and match surplus value V^1. By the same steps as in the proof of Proposition 2, it follows that

$$w_0(s_t,\boldsymbol{\varepsilon}_t) = y_{f_{1t}}(s_t,k_{1t})+\varepsilon_{1t}$$

$$+\frac{\delta}{1-\delta}\int_{\boldsymbol{\varepsilon}_{t+1}}\left[\eta_1(h_t)EV^1(\cdot|s_t,k_{1t})-\eta_0(h_t)EV^w(\cdot|s_t,k_{0t})-\eta_0(h_t)E\Pi^1(\cdot|s_t,k_{0t})\right]dG, \tag{27}$$

where the third term on the right side of (27) is simply $\Psi(s_t,k_{0t})$ by definition of f_{0t} and f_{1t}. Both

(26) and (27) can be written more compactly as

$$w_0(s_t,\boldsymbol{\varepsilon}_t) = y_{f_t}(s_t,k_{f_t t}) + \varepsilon_{f_t t}$$

$$+\frac{\delta}{1-\delta}\int_{\boldsymbol{\varepsilon}_{t+1}}\left[\eta_{f_t}(h_t)EV^{f_t}(\cdot|s_t,k_{f_t t}) - \eta_0(h_t)EV^w(\cdot|s_t,k_{0t}) - \eta_{e_t k_{e_t t}}(h_t)E\Pi^{f_t}(\cdot|s_t,k_{e_t t})\right]dG$$

where $f_t \in \{f_{0t}, f_{1t}\}$ and $e_t \in \{A, f_{0t}\}$ denotes the employing firm at state $(s_t, \boldsymbol{\varepsilon}_t)$.

Now, note that by definition the worker's value from accepting firm f_{0t}'s offer is $\max_{f\neq A} V_f^w(s_t,\boldsymbol{\varepsilon}_t)$. By (13) and (14), when firm A employs the worker,

$$\max_{f\neq A} V_f^w(s_t,\boldsymbol{\varepsilon}_t) = \max_{f\neq A}\left\{\max_{k\in K^f}\left\{(1-\delta)\left[y_f(s_t,k)+\varepsilon_{fkt}\right] + \delta\eta_{fk}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^f(\cdot|s_t,k)dG\right\},\right.$$
$$\left. -\delta\eta_{Ak_{At}}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^f(\cdot|s_t,k_{At})dG\right\} \tag{28}$$

which, by the properties of type I extreme value distributions and the definition of V^A, leads to

$$V^A(s_t,\boldsymbol{\varepsilon}_t,f_{0t}) = \ln\sum\nolimits_{f\neq A,k\in K^f}\exp\left\{(1-\delta)y_f(s_t,k) + \delta\eta_{fk}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^f(\cdot|s_t,k)dG\right.$$
$$\left. -\delta\eta_{Ak_{At}}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^f(\cdot|s_t,k_{At})dG\right\} + \delta\eta_{0t}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^A(\cdot|s_t,k_{0t})dG + (1-\delta)\varepsilon_{0t}. \tag{29}$$

Here ε_{0t} is the type I extreme value shock at the job of the firm offering the second-highest value of wages, firm f_{0t}. A similar logic implies that when firm f_{0t} employs the worker, the worker's value from accepting firm f_{0t}'s offer is $\max_{f\neq f_{0t}} V_f^w(s_t,\boldsymbol{\varepsilon}_t)$, which leads to

$$V^A(s_t,\boldsymbol{\varepsilon}_t,f_{0t}) = \ln\sum\nolimits_{f\neq f_{0t},k\in K^f}\exp\left\{(1-\delta)y_f(s_t,k) + \delta\eta_{fk}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^f(\cdot|s_t,k)dG\right.$$
$$\left. -\delta\eta_{0t}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^f(\cdot|s_t,k_{0t})dG\right\} + \delta\eta_{0t}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^A(\cdot|s_t,k_{0t})dG + (1-\delta)\varepsilon_{1t} \tag{30}$$

by logic similar to the above one. Here ε_{1t} is the type I extreme value shock at the job of the firm offering the second-highest value of wages, firm f_{1t}.

In the oligopoly case in which firms different from A face no cost of technology adoption and their jobs entail the same prospects for learning, human capital acquisition, and exogenous separation, it follows that $\Pi^f = 0$, and (29) and (30) reduce to

$$\max_{f\neq A,k\in K^f}\left\{(1-\delta)[y_f(s_t,k)+\varepsilon_{fkt}] + \delta\eta_{fk}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k)dG\right\} \tag{31}$$

as proved in Proposition 2. By the properties of type I extreme value distributions, $V^A(s_t,\boldsymbol{\varepsilon}_t,f_{0t})$ in

(31) can also be expressed as

$$\ln \sum\nolimits_{f \neq A, k \in K^f} \exp \left\{ (1-\delta) y_f(s_t, k) + \delta \eta_{fk}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1} | s_t, k) dG \right\} + (1-\delta)\varepsilon_{0t},$$

where ε_{0t} is the productivity shock at the job offered by firm f_{0t}.

Lastly, note that, as argued in the proof of Proposition 1, $w_A(s_t, \boldsymbol{\varepsilon}_t)$ must be such that the worker is indifferent between firm A's and firm f_{0t}'s offers. Therefore,

$$w_A(s_t, \boldsymbol{\varepsilon}_t) = \frac{\max_{f \neq A} V_f^w(s_t, \boldsymbol{\varepsilon}_t)}{1-\delta} - \frac{\delta \eta_{Ak_{At}}(h_t)}{1-\delta} \int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(\cdot | s_t, k_{At}) dG,$$

which, by using (28), the properties of type I extreme value distributions, and the definition of V^f, can be rewritten as

$$w_A(s_t, \boldsymbol{\varepsilon}_t) = \ln \sum\nolimits_{f \neq A, k \in K^f} \exp \left\{ y_f(s_t, k) + \frac{\delta \eta_{fk}(h_t)}{1-\delta} \int_{\boldsymbol{\varepsilon}_{t+1}} EV^f(s_{t+1}, \boldsymbol{\varepsilon}_{t+1} | s_t, k) dG \right.$$

$$\left. - \frac{\delta \eta_{Ak_{At}}(h_t)}{1-\delta} \int_{\boldsymbol{\varepsilon}_{t+1}} EV^f(s_{t+1}, \boldsymbol{\varepsilon}_{t+1} | s_t, k_{At}) dG \right\} + \varepsilon_{0t}$$

or, equivalently,

$$w_A(s_t, \boldsymbol{\varepsilon}_t) = \ln \sum\nolimits_{f \neq A, k \in K^f} \exp \left\{ y_f(s_t, k) + \Psi_k(s_t, k_{At}) \right\} + \varepsilon_{0t}, \tag{32}$$

where

$$\Psi_{fk}(s_t, k_{At}) = \frac{\delta \eta_{fk}(h_t)}{1-\delta} \int_{\boldsymbol{\varepsilon}_{t+1}} EV^f(s_{t+1}, \boldsymbol{\varepsilon}_{t+1} | s_t, k) dG - \frac{\delta \eta_{Ak_{At}}(h_t)}{1-\delta} \int_{\boldsymbol{\varepsilon}_{t+1}} EV^f(s_{t+1}, \boldsymbol{\varepsilon}_{t+1} | s_t, k_{At}) dG.$$

With workers of different skill types i (see the paper) and

$$y_f(s_{it}, k) + \varepsilon_{fkt} = \gamma_{fk} + \underbrace{\gamma_{Li} h_t + \beta_{fk} \gamma_{fHi}}_{b(h_t, i)} + \underbrace{(\alpha_{fk} - \beta_{fk}) \gamma_{fHi}}_{c(i)} p_{it} + \varepsilon_{fkt}, \tag{33}$$

derived in the paper in Section 4.2, it follows that (32) for workers continually employed by firm A can be rewritten as

$$w_{Ait} = w_A(s_{it}, \boldsymbol{\varepsilon}_t) = b(h_t, i) + c(i) p_{it} + \ln \sum\nolimits_{g \neq A, k \in K^g} \exp \left\{ \gamma_{gk} + \Psi_k(s_{it}, k_{At}) \right\} + \varepsilon_{0t}, \tag{34}$$

which is the expression for w_{Ait} in Section 4.3 in the paper. I summarize this result in the following:

Corollary 1. *Suppose that $y_{fLk}(h_t, i) = \gamma_{fk} + \gamma_{Li} h_t$ and $y_{fHk}(h_t, i) = y_{fLk}(h_t) + \gamma_{fHi}(t-1)$ at each firm and job. Suppose that for firm A's competitors, $\gamma_{fHi}(t-1)$ is independent of $t-1$ and that $\beta_{fk} \gamma_{fHi}$ and $(\alpha_{fk} - \beta_{fk}) \gamma_{fHi}$, respectively, are constant across firms and jobs. Then, a worker's* *equilibrium wage when employed by firm A at job k_{At} is*

$$w_{Ait} = w_A(s_{it}, \boldsymbol{\varepsilon}_t) = b(h_t, i) + c(i)p_{it} + \ln \sum\nolimits_{g \neq A, k \in K^g} \exp\left\{\gamma_{gk} + \Psi_{gk}(s_{it}, k_{At})\right\} + \varepsilon_{0t}. \tag{35}$$

I turn now to characterize firm A's match surplus value in the oligopoly case considered in the paper.

Proposition 3. *Under the assumption that for any technology used by a competitor of firm A, at least one other firm has access to the same technology, the wage offered by any such firm $f \neq A$ is $w_f(s_t, \boldsymbol{\varepsilon}_t) = y_f(s_t, k_{ft}) + \varepsilon_{fk_{ft}t}$, where $k_{ft} = k_f(s_t, \boldsymbol{\varepsilon}_t)$ is the job offered by any firm $f \neq A$. If, in addition, the technologies of all these firms entail the same prospects for learning, human capital acquisition, and exogenous separation, then firm A's match surplus value in (14) reduces to*

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

$$\left.\max_{f \neq A, k \in K^f}\left\{(1-\delta)[y_f(s_t, k) + \varepsilon_{fkt}] + \delta\eta_{fk}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k)dG\right\}\right\} \tag{36}$$

and $y_{f_{0t}}(s_t, k_{0t}) + \varepsilon_{0t} = \max_{f \neq A, k \in K^f}\{y_f(s_t, k) + \varepsilon_{fkt}\}$.

Proof. Since, by assumption, any competitor of firm A faces competition, among others, from a firm operating its same technology, any such firm makes zero profit, as argued in the paper. In particular, the continuation profits of any such firm are zero, so

$$\Pi^f(s_t, \boldsymbol{\varepsilon}_t) = \max_{w,k} d_{ft}\left\{(1-\delta)[y_f(s_t, k) + \varepsilon_{fkt} - w]\right\} = 0 \tag{37}$$

where $d_{ft} = d_f(s_t, \boldsymbol{\varepsilon}_t, \mathbf{w}_t, \mathbf{k}_t)$ is the acceptance decision of the worker with respect to firm f's offer. These observations imply that the wage offered by any such firm must equal its expected output (including the realized productivity shock) at the offered job. In particular, this is true for firm f_{0t}, whose job offer is k_{0t}. Thus, the wage offer of firm f_{0t} is $w_0(s_t, \boldsymbol{\varepsilon}_t) = y_{f_{0t}}(s_t, k_{0t}) + \varepsilon_{0t}$ and (14) can be rewritten as

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

$$\left.(1-\delta)\left[y_{f_{0t}}(s_t, k_{0t}) + \varepsilon_{0t}\right] + \delta\eta_0(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k_{0t})dG\right\}. \tag{38}$$

I now prove that if the technologies of all firms different from A entail the same prospects for learning, human capital acquisition, and exogenous separation, the second branch of (38) reduces to

$$\max_{f \neq A, k \in K^f}\left\{(1-\delta)[y_f(s_t, k) + \varepsilon_{fkt}] + \delta\eta_{fk}(h_t)\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t, k)dG\right\}.$$

To see this, note first that equilibrium, in particular (13), implies that the job choice of each firm $f \neq A$, job k_{ft}, solves the relevant version of (15), that is,

$$\max_{k \in K^f} \left\{ (1-\delta)[y_f(s_t,k)+\varepsilon_{fkt}] + \delta\eta_{fk}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k)dG \right\}.$$

So, by (13), the match surplus value of any firm $f \neq A$ is given by

$$V^f(s_t,\boldsymbol{\varepsilon}_t) = \max_{k \in K^f} \left\{ (1-\delta)[y_f(s_t,k)+\varepsilon_{fkt}] + \delta\eta_{fk}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k)dG \right\}. \tag{39}$$

Since no such firm f makes a profit, $V^f = V^w$. Hence, the competitor of firm A most preferred by the worker, firm f_{0t}, is the firm with the highest match surplus value, that is, the firm with $V^0(s_t,\boldsymbol{\varepsilon}_t) = \max_{f \neq A}\{V^f(s_t,\boldsymbol{\varepsilon}_t)\}$ or, equivalently, the firm offering the worker the following (expected present discounted) value of wages,

$$\max_{f \neq A, k \in K^f} \left\{ (1-\delta)[y_f(s_t,k)+\varepsilon_{fkt}] + \delta\eta_{fk}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k)dG \right\} \tag{40}$$

given that $V^f = V^w$. Since $E\Pi^A(\cdot|s_t,k_{ft})$ is independent of k_{ft} for any $f \neq A$, $V^A = V^w + \Pi^A$, when the worker is not employed by firm A, can be equivalently expressed as

$$\max_{f \neq A, k \in K^f} \Bigg\{ (1-\delta)[y_f(s_t,k)+\varepsilon_{fkt}] + \delta\eta_{fk}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k)dG$$

$$+\delta\eta_{fk}(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} E\Pi^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k)dG \Bigg\}.$$

Given that $V^A = \Pi^A + V^w$, this completes the proof of the first part of the claim.

As for the second part of the claim, note that, by assumption, $EV^w(\cdot|s_t,k_{ft})$ is independent of k_{ft} for any $f \neq A$. Hence, (40) can also be rewritten as

$$(1-\delta) \max_{f \neq A, k \in K^f} \{y_f(s_t,k)+\varepsilon_{fkt}\} + \delta\eta_0(h_t) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{0t})dG$$

so the expected output $y_{f_{0t}}(s_t,k_{0t})+\varepsilon_{0t}$ that the worker would produce at firm f_{0t} is equal to

$$\max_{f \neq A, k \in K^f} \{y_f(s_t,k)+\varepsilon_{fkt}\} = \max\nolimits_{f \neq A, k \in K^f} \{y_f(s_t,k)+\varepsilon_{fkt}\},$$

which also equals the wage offered by firm f_{0t} since, as argued above, $w_0(s_t,\boldsymbol{\varepsilon}_t) = y_{f_{0t}}(s_t,k_{0t})+\varepsilon_{0t}$. This completes the proof of the claim. □

2 Identification

The model is fit to a rich panel dataset on the sequence of yearly job assignments, paid wages, and recorded performance ratings for more than 1,400 managers over eight years, in the case of the baseline sample, and for more than 2,000 managers over eight years, in the case of the extended sample. (See the estimation exercise described in Section 5.) In this Section I establish the (nonparametric and parametric) identification of the model based on these data. The Monte Carlo results reported in Section 4.3 provide additional evidence on the fact that the rich information in the data is sufficient to pin down quite precisely the model's parameters.

2.1 Discrete Choice Component of the Model: Employment and Job Assignment

Consider the employment and job assignment problem of my firm over the first eight years of tenure of a manager, which constitutes the discrete choice component of my model. In the model, employment and job assignment depend on the beliefs about a manager's ability, a manager's human capital, and the realization of (type I extreme value) productivity shocks. Recall that initial prior beliefs, which are unobserved by the econometrician, are modeled as a nonparametric finite mixture distribution.[2] Hence, the discrete choice component of the model is a nonparametric mixture model of parametric (type I extreme value) component distributions. The identification of the discrete choice component of the model amounts to the identification of the process for the observed assignment k_t, the unobserved prior p_t, and the observed human capital h_t.

Now, by the law of conditional probability, we know that

$$\Pr(k_{t+1}, p_{t+1}, h_{t+1}|p_t, h_t, k_t) = \Pr(k_{t+1}|p_t, h_t, k_t, p_{t+1}, h_{t+1}) \Pr(h_{t+1}|p_t, h_t, k_t, p_{t+1}) \Pr(p_{t+1}|p_t, h_t, k_t), \tag{41}$$

which, using the implications of the model, can be simplified to

$$\Pr(k_{t+1}, p_{t+1}, h_{t+1}|p_t, h_t, k_t) = \Pr(k_{t+1}|p_{t+1}, h_{t+1}) \Pr(h_{t+1}|h_t, k_t) \Pr(p_{t+1}|p_t, k_t). \tag{42}$$

The equality of the first term on the right side of (41) with the first term on the right side of (42) follows because the next period job assignment of a manager depends only on the next period prior and human capital by the Markovian nature of the match surplus value problem of a firm. The equality of the second terms in these expressions follows because the next period human capital depends only on the current period human capital and job assignment by assumption (see the specification of the process of human capital acquisition in the paper). The equality of the third terms follows because the next period prior depends only on the current period prior and job assignment by Bayes' rule.

Observe that the process $\Pr(p_{t+1}|p_t, k_t)$ for the prior does not depend on either a manager's unobserved ability θ or a manager's skill type i. That this law of motion does not depend on θ is immediately implied by the model, because θ is unknown to all model agents. That the law of motion $\Pr(p_{t+1}|p_t, k_t)$

[2] In the spirit of the test by Pakes and Ericson (1998) for Bayesian learning, the fact that the empirical process for observed performance ratings, as well as the empirical process for job assignments and wages, does not appear to be first-order Markov provides evidence for the presence of learning.

does not depend on a manager's skill type i follows because the parameters $\{\alpha_k, \beta_k\}_{k=1}^3$ governing the distribution of true performance signals about ability, and hence the law of motion for beliefs, are assumed to be independent of a manager's skill type. Technically, this feature of the discrete choice component of the model rules out serially correlated individual-specific heterogeneity in employment and job assignment conditional on p_t. In particular, if the distribution of true performance is known, then the only unknown component of $\Pr(p_{t+1}|p_t, k_t)$ is the distribution $\Pr(p_1|p_0, k_0) \equiv \Pr(p_1)$ of 'initial conditions'. Thus, the identification problem for the law of motion for p_t reduces to the problem of identifying the distribution of initial priors, $\Pr(p_1)$, which I specify as a finite mixture with known components, $i = 1, \ldots, I$.[3]

Below I first establish the identification of the learning process determining the evolution of beliefs based on information on managers' performance ratings. I then establish the identification of the output and human capital process, as well as of the job assignment process. For simplicity, I omit the subscripts A for the firm and n for a manager whenever unambiguous.

2.1.1 Identification of the Learning Process

Here I argue the identification of the process for $\Pr(p_{t+1}|p_t, k_t)$ conditional on $\Pr(p_1)$. Specifically, I show that the observed distribution of performance ratings provides a direct source of identification for $\{\alpha_k, \beta_k\}_{k=1}^3$ and the parameters governing classification error in reported ratings.

I divide the argument into two cases. In the first case, *Case 1*, I prove that repeated observations on performance ratings allow me straightforwardly to identify $\{\alpha_k, \beta_k\}_{k=1}^3$ in the absence of classification error. In the second case, *Case 2*, I show that identification can be established also in the presence of classification error by a simple extension of the argument in Case 1. This augmented argument proves that $\{\alpha_k, \beta_k\}_{k=1}^3$ and the classification error parameters are identified.

Case 1: Argument without Classification Error. I will show here that three periods of observations on performance at Level 1 allow me to identify α_1, β_1, and the initial prior (or the mean initial prior, in the presence of multiple skill types). Suppose, first, for simplicity, that all managers are of the same skill type. Consider managers retained at my firm, firm A, at Level 1 for at least three years. Let m^o_{Akt} denote the observed proportion of high ratings at Level k of my firm in period t. Assume that $m^o_{Akt} > 0$ at all k and t. By equating the sample proportion of individuals with one high rating at the end of period 1, denoted by m^o_{A11}, with another high rating at the end of period 2, denoted by $m^o_{A12|H1}$, and with yet another high rating at the end of period 3, denoted by $m^o_{A13|HH1}$, respectively, to their theoretical counterparts, I obtain

$$\begin{cases} \beta_1 + (\alpha_1 - \beta_1)p_1 = m^o_{A11} \\ \dfrac{\alpha_1^2 p_1}{\alpha_1 p_1 + \beta_1(1-p_1)} + \dfrac{\beta_1^2(1-p_1)}{\alpha_1 p_1 + \beta_1(1-p_1)} = m^o_{A12|H1} \Rightarrow \beta_1^2 + (\alpha_1^2 - \beta_1^2)p_1 = m^o_{A12|H1} m^o_{A11} \\ \dfrac{\alpha_1^3 p_1}{\alpha_1^2 p_1 + \beta_1^2(1-p_1)} + \dfrac{\beta_1^3(1-p_1)}{\alpha_1^2 p_1 + \beta_1^2(1-p_1)} = m^o_{A13|HH1} \Rightarrow \beta_1^3 + (\alpha_1^3 - \beta_1^3)p_1 = m^o_{A13|HH1} m^o_{A12|H1} m^o_{A11} \end{cases}$$

[3] Notice the usual lack of identification with respect to i, because the type distribution is invariant to permutations of the points in its support. See Buchinsky, Hahn, and Kim (2010) for caveats regarding the identifiability of finite mixtures with known components in applied frameworks.

by using the definition of $P_{AH1}(p_1) = \alpha_1 p_1/[\alpha_1 p_1 + \beta_1(1-p_1)]$ and $P^2_{AH1}(p_1) = \alpha_1^2 p_1/[\alpha_1^2 p_1 + \beta_1^2(1-p_1)]$. I will now show that this system identifies α_1, β_1, and p_1. To this purpose, note that the three equations in the above system can also be expressed as

$$\begin{cases} (\alpha_1 - \beta_1)p_1 = m^o_{A11} - \beta_1 \\ (\alpha_1 - \beta_1)(\alpha_1 + \beta_1)p_1 = m^o_{A12|H1} m^o_{A11} - \beta_1^2 \\ (\alpha_1 - \beta_1)(\alpha_1^2 + \alpha_1\beta_1 + \beta_1^2)p_1 = m^o_{A13|HH1} m^o_{A12|H1} m^o_{A11} - \beta_1^3 \end{cases} \tag{43}$$

or, using the fact that $(\alpha_1 - \beta_1)p_1 = m^o_{A11} - \beta_1$ and $\alpha_1^2 + \alpha_1\beta_1 + \beta_1^2 = (\alpha_1 + \beta_1)^2 - \alpha_1\beta_1$,

$$\begin{cases} (\alpha_1 - \beta_1)p_1 = m^o_{A11} - \beta_1 \\ (m^o_{A11} - \beta_1)(\alpha_1 + \beta_1) = m^o_{A12|H1} m^o_{A11} - \beta_1^2 \\ (m^o_{A12|H1} m^o_{A11} - \beta_1^2)(\alpha_1 + \beta_1) - \alpha_1\beta_1 \, (m^o_{A11} - \beta_1) = m^o_{A13|HH1} m^o_{A12|H1} m^o_{A11} - \beta_1^3 \end{cases} . \tag{44}$$

Rewriting the third equality in (44), I obtain

$$m^o_{A12|H1} m^o_{A11} \, (\alpha_1 + \beta_1) - m^o_{A11}\alpha_1\beta_1 = m^o_{A13|HH1} m^o_{A12|H1} m^o_{A11}$$

or, equivalently, if $m^o_{A11} > 0$,

$$m^o_{A12|H1}(\alpha_1 + \beta_1) - \alpha_1\beta_1 = m^o_{A13|HH1} m^o_{A12|H1} \tag{45}$$

whereas rewriting the second equality in (44), I obtain

$$m^o_{A11}(\alpha_1 + \beta_1) - \alpha_1\beta_1 = m^o_{A12|H1} m^o_{A11}. \tag{46}$$

By combining (45) and (46) into a system, it follows

$$\begin{cases} m^o_{A12|H1}(\alpha_1 + \beta_1) - \alpha_1\beta_1 = m^o_{A13|HH1} m^o_{A12|H1} \\ m^o_{A11}(\alpha_1 + \beta_1) - \alpha_1\beta_1 = m^o_{A12|H1} m^o_{A11} \end{cases} \tag{47}$$

and taking the difference between the two expressions in (47) side by side leads to

$$\alpha_1 + \beta_1 = \frac{m^o_{A12|H1}\left(m^o_{A13|HH1} - m^o_{A11}\right)}{m^o_{A12|H1} - m^o_{A11}} \tag{48}$$

where $m^o_{A13|HH1} \geq m^o_{A12|H1} \geq m^o_{A11}$ if $P^2_{AH1}(p_1) \geq P_{AH1}(p_1) \geq p_1$, which is implied by $\alpha_1 \geq \beta_1$. Substituting the expression in (48) into the second expression in (47), I obtain

$$\alpha_1\beta_1 = \frac{m^o_{A11} m^o_{A12|H1}\left(m^o_{A13|HH1} - m^o_{A11}\right)}{m^o_{A12|H1} - m^o_{A11}} - m^o_{A12|H1} m^o_{A11} = \frac{m^o_{A11} m^o_{A12|H1}\left(m^o_{A13|HH1} - m^o_{A12|H1}\right)}{m^o_{A12|H1} - m^o_{A11}}. \tag{49}$$

Now, using the fact that $(\alpha_1-\beta_1)^2 = (\alpha_1+\beta_1)^2 - 4\alpha_1\beta_1$, by (48) and (49) it follows

$$(\alpha_1-\beta_1)^2 = \left(m^o_{A12|H1}\right)^2 \left(\frac{m^o_{A13|HH1} - m^o_{A11}}{m^o_{A12|H1} - m^o_{A11}}\right)^2 \left[1 - \frac{4m^o_{A11}\left(m^o_{A13|HH1} - m^o_{A12|H1}\right)\left(m^o_{A12|H1} - m^o_{A11}\right)}{m^o_{A12|H1}\left(m^o_{A13|HH1} - m^o_{A11}\right)^2}\right].$$

Hence, the fact that $\alpha_1 \geq \beta_1$ leads to

$$\alpha_1 - \beta_1 = m^o_{A12|H1}\left(\frac{m^o_{A13|HH1} - m^o_{A11}}{m^o_{A12|H1} - m^o_{A11}}\right)\sqrt{1 - \frac{4m^o_{A11}\left(m^o_{A13|HH1} - m^o_{A12|H1}\right)\left(m^o_{A12|H1} - m^o_{A11}\right)}{m^o_{A12|H1}\left(m^o_{A13|HH1} - m^o_{A11}\right)^2}}. \tag{50}$$

Summing (48) and (50), I obtain

$$2\alpha_1 = \frac{m^o_{A12|H1}\left(m^o_{A13|HH1} - m^o_{A11}\right)}{m^o_{A12|H1} - m^o_{A11}}$$

$$+m^o_{A12|H1}\left(\frac{m^o_{A13|HH1} - m^o_{A11}}{m^o_{A12|H1} - m^o_{A11}}\right)\sqrt{1 - \frac{4m^o_{A11}\left(m^o_{A13|HH1} - m^o_{A12|H1}\right)\left(m^o_{A12|H1} - m^o_{A11}\right)}{m^o_{A12|H1}\left(m^o_{A13|HH1} - m^o_{A11}\right)^2}}$$

so α_1 is identified. Then, β_1 is also identified. Note that in this case p_1 is identified too by $(\alpha_1-\beta_1)p_1 = m^o_{A11} - \beta_1$. By repeating this argument, with α_1, β_1, and p_1 known, for managers promoted from Level 1 to Level 2, and continually assigned to Level 2 for at least two periods, and then for managers promoted from Level 2 to Level 3, and continually assigned to Level 3 for at least two periods, it is also possible to establish identification of α_2, β_2, α_3, and β_3. If managers were of different skill types, then an analogous argument would apply. Specifically, in the above expressions p_1 would be replaced by the mean initial prior over the skill types of managers who are retained at Level 1 for at least three years. So, this mean initial prior and the learning parameters would still be identified.

Observe also that higher moments of the distribution of performance ratings provide additional identifying restrictions. To see this, consider the variance of the first period ratings at Level 1, which can be expressed as

$$\alpha_1 p_1 + \beta_1(1-p_1) - [\alpha_1 p_1 + \beta_1(1-p_1)]^2 = m^o_{A11} - m^o_{A12|H1}m^o_{A11} + (\alpha_1-\beta_1)^2 p_1(1-p_1) \tag{51}$$

since $\beta_1^2 + (\alpha_1^2 - \beta_1^2)p_1 = m^o_{A12|H1}m^o_{A11}$. Denoting the empirical variance of the first period ratings by $\sigma^o_{r_{A11}}$, from (51) it follows

$$(\alpha_1-\beta_1)^2 p_1(1-p_1) = \sigma^o_{r_{A11}} - m^o_{A11} + m^o_{A12|H1}m^o_{A11}.$$

As for the panel length required for the identification of the learning parameters $\{\alpha_k, \beta_k\}_{k=1}^3$, note that with p_1 known, the two conditions

$$\begin{cases} \beta_1 + (\alpha_1 - \beta_1)\, p_1 = m^o_{A11} \\ \beta_1^2 + \left(\alpha_1^2 - \beta_1^2\right) p_1 = m^o_{A12} \end{cases}, \tag{52}$$

where $m^o_{A12} = m^o_{A12|H1} m^o_{A11}$, already pin down α_1 and β_1. To see this, observe that

$$\begin{cases} \beta_1 = \left(m^o_{A11} - \alpha_1 p_1\right) / \left(1 - p_1\right) \\ \beta_1^2 + \left(\alpha_1^2 - \beta_1^2\right) p_1 = m^o_{A12} \Rightarrow \alpha_1^2 p_1 + \frac{\left(m^o_{A11} - \alpha_1 p_1\right)^2}{(1-p_1)^2} (1 - p_1) - m^o_{A12} = 0 \end{cases}$$

so the second condition in the above system can be rewritten as

$$\alpha_1^2 p_1 - 2 m^o_{A11} \alpha_1 p_1 + \left(m^o_{A11}\right)^2 - (1 - p_1)\, m^o_{A12} = 0$$

with solutions

$$\begin{aligned} \alpha_1 &= \frac{2 m^o_{A11} p_1 \pm \sqrt{4\left(m^o_{A11}\right)^2 p_1^2 - 4 p_1 \left[\left(m^o_{A11}\right)^2 - (1 - p_1)\, m^o_{A12}\right]}}{2 p_1} \\ &= m^o_{A11} \pm \sqrt{(1 - p_1)\left[m^o_{A12} - \left(m^o_{A11}\right)^2\right] / p_1} \end{aligned} \tag{53}$$

where $m^o_{A12} \geq \left(m^o_{A11}\right)^2$ given that $m^o_{A12} - \left(m^o_{A11}\right)^2 = p_1 (1 - p_1)(\alpha_1 - \beta_1)^2 \geq 0$ (with strict inequality for $p_1 \in (0, 1)$ and $\alpha_1 > \beta_1$). But, if $\alpha_1 = m^o_{A11} - \sqrt{(1 - p_1)\left[m^o_{A12} - \left(m^o_{A11}\right)^2\right] / p_1}$, then

$$\alpha_1 < \beta_1 = m^o_{A11} + \sqrt{p_1 \left[m^o_{A12} - \left(m^o_{A11}\right)^2\right] / (1 - p_1)},$$

which is inconsistent with the assumptions of the model. Plugging this expression for α_1 back into $\beta_1 = \left(m^o_{A11} - \alpha_1 p_1\right) / (1 - p_1)$, I obtain

$$\beta_1 = m^o_{A11} - \sqrt{p_1 \left[m^o_{A12} - \left(m^o_{A11}\right)^2\right] / (1 - p_1)}.$$

Hence, given $p_1 \in (0, 1)$ and $\alpha_1 > \beta_1$, the system in (52) is sufficient to recover α_1 and β_1. An analogous argument can be used to show that just two periods of observation of performance at Levels 2 and 3 are sufficient to identify (α_2, β_2) and (α_3, β_3), respectively.

Case 2: Argument with Classification Error. I will show here that repeated observations on the performance ratings of managers continually assigned to a level allow me to pin down the parameters of the classification error in recorded performance in addition to the learning parameters $\{\alpha_k, \beta_k\}_{k=1}^3$. To start, first ignore skill types for simplicity. Consider managers continually assigned to Level 1 up to tenure t. Note that, in the presence of classification error as specified in the paper, the probability of an observed high rating is $E_0(L1, t) + [1 - E_1(L1, t) - E_0(L1, t)]\, [\beta_1 + (\alpha_1 - \beta_1) p_1]$. Equating this probability to its sample counterpart, m^o_{A1t}, I obtain

$$E_0(L1,t) + [1 - E_1(L1,t) - E_0(L1,t)]\,[\beta_1 + (\alpha_1 - \beta_1)p_1] = m^o_{A1t},$$

which, letting $\nu = \beta_1 + (\alpha_1 - \beta_1)p_1$, can be rewritten as

$$m^o_{A1t} = (1-\nu)\underbrace{\frac{\exp\{d_0 + d_2(L1)t\}}{1+\exp\{d_0 + d_2(L1)t\}}}_{E_0(L1,t)} + \nu\underbrace{\frac{\exp\{d_0 + d_1 + d_2(L1)t\}}{1+\exp\{d_0 + d_1 + d_2(L1)t\}}}_{1-E_1(L1,t)}.$$

Hence, the population distribution of performance ratings can be thought as the nonparametric mixture of two logistic models with the same coefficient on the explanatory variable, here tenure. So, ν, d_0, d_1, and $d_2(L1)$ are identified. Now, if $\nu = \beta_1 + (\alpha_1 - \beta_1)p_1$ is known, then $\beta_1 = (\nu - \alpha_1 p_1)\,/\,(1-p_1)$ is also known up to α_1 and p_1. Observe, for later, that ν is the first population moment of the true distribution of performance, that is, $\nu = m_{A11}$. (Moments of the form m_{Akt} and similar are moments of the true rather than of the observed distribution of performance ratings.)

I now prove that α_1 is also identified (as well as p_1). First, note that by the law of conditional probability, we have

$$\Pr(R^o_2 = 1|R_1 = 1) = \Pr(R^o_2 = 1|R_1 = 1, R_2 = 1)\Pr(R_2 = 1|R_1 = 1)$$

$$+\Pr(R^o_2 = 1|R_1 = 1, R_2 = 0)\Pr(R_2 = 0|R_1 = 1).$$

Now, using the fact that the error in the recording rating in a period, conditional on the true rating, is independent of past true ratings, which implies $\Pr(R^o_2 = 1|R_1 = 1, R_2 = 1) = \Pr(R^o_2 = 1|R_2 = 1)$, it follows that

$$\Pr(R^o_2 = 1|R_1 = 1) = E_0(L1,2) + [1 - E_0(L1,2) - E_1(L1,2)]\left[\frac{\alpha_1^2 p_1 + \beta_1^2(1-p_1)}{\alpha_1 p_1 + \beta_1(1-p_1)}\right].$$

Observe that

$$\Pr(R^o_2 = 1|R^o_1 = 1) = \Pr(R^o_2 = 1|R^o_1 = 1, R_2 = 1)\Pr(R_2 = 1|R^o_1 = 1)$$

$$+\Pr(R^o_2 = 1|R^o_1 = 1, R_2 = 0)\Pr(R_2 = 0|R^o_1 = 1)$$

$$= \Pr(R^o_2 = 1|R_2 = 0) + [\Pr(R^o_2 = 1|R_2 = 1) - \Pr(R^o_2 = 1|R_2 = 0)]\Pr(R_2 = 1|R^o_1 = 1)$$

$$= E_0(L1,2) + [1 - E_0(L1,2) - E_1(L1,2)]\Pr(R_2 = 1|R^o_1 = 1), \qquad (54)$$

where I have used again the fact that the error in the recording rating in a period, conditional on the true rating, is independent of past recorded ratings, so $\Pr(R^o_2 = 1|R^o_1 = 1, R_2 = r) = \Pr(R^o_2 = 1|R_2 = r)$, $r \in \{0,1\}$. Next, note that

$$\Pr(R_2 = 1|R^o_1 = 1) = \frac{\Pr(R_2 = 1, R^o_1 = 1)}{\Pr(R^o_1 = 1)} = \frac{\Pr(R_2 = 1|R^o_1 = 1, R_1 = 1)\Pr(R^o_1 = 1|R_1 = 1)\Pr(R_1 = 1)}{\Pr(R^o_1 = 1)}$$

$$+\frac{\Pr(R_2=1|R_1^o=1,R_1=0)\Pr(R_1^o=1|R_1=0)\Pr(R_1=0)}{\Pr(R_1^o=1)}$$

so with $\Pr(R_1^o=1)=m^o_{A11}$, $\Pr(R_1=1)=\nu$, $\Pr(R_1^o=1|R_1=0)=E_0(L1,1)$, and $\Pr(R_1^o=1|R_1=1)=1-E_1(L1,1)$ by definition, we have

$$\Pr(R_2=1|R_1^o=1)$$

$$=\frac{\left[\alpha_1^2p_1+\beta_1^2(1-p_1)\right][1-E_1(L1,1)]+[\alpha_1(1-\alpha_1)p_1+\beta_1(1-\beta_1)(1-p_1)]\,E_0(L1,1)}{m^o_{A11}}. \tag{55}$$

Also by definition, $\Pr(R_2^o=1|R_1^o=1)=m^o_{A12|H1}$ and $\Pr(R_2=1|R_1=1)=m_{A12|H1}$. Then, I can rewrite (54) as

$$\Pr(R_2=1|R_1^o=1)=\frac{m^o_{A12|H1}-E_0(L1,2)}{1-E_0(L1,2)-E_1(L1,2)}$$

or, equivalently, by using (55),

$$\alpha_1^2p_1+\beta_1^2(1-p_1)=\frac{1}{1-E_0(L1,1)-E_1(L1,1)}\left[\frac{m^o_{A11}m^o_{A12|H1}-m^o_{A11}E_0(L1,2)}{1-E_0(L1,2)-E_1(L1,2)}-\nu E_0(L1,1)\right]. \tag{56}$$

Repeat now the same argument for the second year of tenure. Specifically, start from

$$\Pr(R_3^o=1|R_1^o=1,R_2^o=1)=\Pr(R_3^o=1|R_1^o=1,R_2^o=1,R_3=1)\Pr(R_3=1|R_1^o=1,R_2^o=1)$$

$$+\Pr(R_3^o=1|R_1^o=1,R_2^o=1,R_3=0)\Pr(R_3=0|R_1^o=1,R_2^o=1),$$

which can equivalently be expressed as

$$\underbrace{\Pr(R_3^o=1|R_1^o=1,R_2^o=1)}_{m^o_{A13|HH1}}=\underbrace{\Pr(R_3^o=1|R_3=0)}_{E_0(L1,3)}$$

$$+\left[\underbrace{\Pr(R_3^o=1|R_3=1)}_{1-E_1(L1,3)}-\underbrace{\Pr(R_3^o=1|R_3=0)}_{E_0(L1,3)}\right]\Pr(R_3=1|R_1^o=1,R_2^o=1). \tag{57}$$

By repeating the above steps, observe that

$$\Pr(R_3=1|R_1^o=1,R_2^o=1)=\frac{\Pr(R_3=1,R_1^o=1,R_2^o=1)}{\Pr(R_1^o=1,R_2^o=1)}$$

$$=\frac{1}{\Pr(R_1^o=1,R_2^o=1)}\left[\Pr(R_3=1,R_1^o=1,R_2^o=1|R_1=1,R_2=1)\Pr(R_1=1,R_2=1)\right.$$

$$+\Pr(R_3=1,R_1^o=1,R_2^o=1|R_1=1,R_2=0)\Pr(R_1=1,R_2=0)$$

$$+\Pr(R_3=1,R_1^o=1,R_2^o=1|R_1=0,R_2=1)\Pr(R_1=0,R_2=1)$$

$$\left.+\Pr(R_3=1,R_1^o=1,R_2^o=1|R_1=0,R_2=0)\Pr(R_1=0,R_2=0)\right]. \tag{58}$$

Now, note that

$$\Pr(R_3 = 1, R_1^o = 1, R_2^o = 1 | R_1 = 1, R_2 = 1) = \Pr(R_3 = 1 | R_1^o = 1, R_2^o = 1, R_1 = 1, R_2 = 1)$$

$$\cdot \Pr(R_1^o = 1, R_2^o = 1 | R_1 = 1, R_2 = 1)$$

$$= \Pr(R_3 = 1 | R_1 = 1, R_2 = 1) \Pr(R_2^o = 1 | R_2 = 1) \Pr(R_1^o = 1 | R_1 = 1)$$

since $\Pr(R_3 = 1 | R_1^o = 1, R_2^o = 1, R_1 = 1, R_2 = 1)$ is independent of R_1^o and R_2^o, and

$$\Pr(R_1^o = 1, R_2^o = 1 | R_1 = 1, R_2 = 1) = \frac{\Pr(R_1 = 1, R_1^o = 1, R_2 = 1, R_2^o = 1)}{\Pr(R_1 = 1, R_2 = 1)}$$

$$= \frac{\Pr(R_2^o = 1 | R_1 = 1, R_1^o = 1, R_2 = 1) \Pr(R_2 = 1 | R_1 = 1, R_1^o = 1) \Pr(R_1 = 1, R_1^o = 1)}{\Pr(R_1 = 1, R_2 = 1)}$$

$$= \frac{\Pr(R_2^o = 1 | R_2 = 1) \Pr(R_2 = 1 | R_1 = 1) \Pr(R_1^o = 1 | R_1 = 1) \Pr(R_1 = 1)}{\Pr(R_1 = 1, R_2 = 1)}$$

$$= \Pr(R_2^o = 1 | R_2 = 1) \Pr(R_1^o = 1 | R_1 = 1).$$

More generally,

$$\Pr(R_3 = r_3, R_1^o = r_1^o, R_2^o = r_2^o | R_1 = r_1, R_2 = r_2) = \Pr(R_3 = r_3 | R_1 = r_1, R_2 = r_2)$$

$$\cdot \Pr(R_2^o = r_2^o | R_2 = r_2) \Pr(R_1^o = r_1^o | R_1 = r_1)$$

for $r_t \in \{0, 1\}$, $t = 1, 2, 3$, and $r_t^o \in \{0, 1\}$, $t = 1, 2$. Hence, I can rewrite (58) as

$$\Pr(R_3 = 1 | R_1^o = 1, R_2^o = 1) = \frac{1}{\Pr(R_1^o = 1, R_2^o = 1)}$$

$$\cdot [\Pr(R_3 = 1 | R_1 = 1, R_2 = 1) \Pr(R_2^o = 1 | R_2 = 1) \Pr(R_1^o = 1 | R_1 = 1) \Pr(R_1 = 1, R_2 = 1)$$

$$+ \Pr(R_3 = 1 | R_1 = 1, R_2 = 0) \Pr(R_2^o = 1 | R_2 = 0) \Pr(R_1^o = 1 | R_1 = 1) \Pr(R_1 = 1, R_2 = 0)$$

$$+ \Pr(R_3 = 1 | R_1 = 0, R_2 = 1) \Pr(R_2^o = 1 | R_2 = 1) \Pr(R_1^o = 1 | R_1 = 0) \Pr(R_1 = 0, R_2 = 1)$$

$$+ \Pr(R_3 = 1 | R_1 = 0, R_2 = 0) \Pr(R_2^o = 1 | R_2 = 0) \Pr(R_1^o = 1 | R_1 = 0) \Pr(R_1 = 0, R_2 = 0)]$$

or, equivalently,

$$\Pr(R_3 = 1 | R_1^o = 1, R_2^o = 1) = \frac{1}{\Pr(R_1^o = 1, R_2^o = 1)} \left\{ \left[\alpha_1^3 p_1 + \beta_1^3 (1 - p_1) \right] [1 - E_1(L1, 2)] [1 - E_1(L1, 1)] \right.$$

$$+ \left[\alpha_1^2 p_1 + \beta_1^2 (1 - p_1) - \alpha_1^3 p_1 - \beta_1^3 (1 - p_1) \right] E_0(L1, 2) [1 - E_1(L1, 1)]$$

$$+ \left[\alpha_1^2 p_1 + \beta_1^2 (1 - p_1) - \alpha_1^3 p_1 - \beta_1^3 (1 - p_1) \right] [1 - E_1(L1, 2)] E_0(L1, 1)$$

$$\left. + \left[\alpha_1 (1 - 2\alpha_1 + \alpha_1^2) p_1 + \beta_1 (1 - 2\beta_1 + \beta_1^2)(1 - p_1) \right] E_0(L1, 2) E_0(L1, 1) \right\},$$

which can be simplified to

$$\begin{aligned} &\Pr(R_3 = 1|R_1^o = 1, R_2^o = 1) \\ = &\frac{\left[\alpha_1^3 p_1 + \beta_1^3(1-p_1)\right]\left[1 - E_0(L1,1) - E_1(L1,1)\right]\left[1 - E_0(L1,2) - E_1(L1,2)\right] + C}{\Pr(R_1^o = 1, R_2^o = 1)} \end{aligned} \tag{59}$$

with

$$\begin{aligned} C = &\nu E_0(L1,1)E_0(L1,2) \\ &+ m_{A12}\left\{E_0(L1,2)\left[1 - E_0(L1,1) - E_1(L1,1)\right] + E_0(L1,1)\left[1 - E_0(L1,2) - E_1(L1,2)\right]\right\} \end{aligned}$$

and m_{A12} given by (56), since $\alpha_1^2 p_1 + \beta_1^2(1-p_1) = m_{A12}$ by definition of m_{A12}. Finally, I can rewrite (57) as

$$\Pr(R_3 = 1|R_1^o = 1, R_2^o = 1) = \frac{m_{A13|HH1}^o - E_0(L1,3)}{1 - E_0(L1,3) - E_1(L1,3)}$$

or, equivalently, using (59) and the fact that $\Pr(R_1^o = 1, R_2^o = 1) = m_{A12}^o$ by definition, as

$$\begin{aligned} \alpha_1^3 p_1 + \beta_1^3(1-p_1) = &\frac{1}{\left[1 - E_0(L1,1) - E_1(L1,1)\right]\left[1 - E_0(L1,2) - E_1(L1,2)\right]} \\ &\cdot \left[\frac{m_{A12}^o m_{A13|HH1}^o - m_{A12}^o E_0(L1,3)}{1 - E_0(L1,3) - E_1(L1,3)} - C\right]. \end{aligned} \tag{60}$$

Lastly, with $\nu = m_{A11}$, (56), and (60), I obtain the system

$$\begin{cases} \beta_1 + (\alpha_1 - \beta_1)\, p_1 = \nu \\ \beta_1^2 + \left(\alpha_1^2 - \beta_1^2\right) p_1 = \frac{1}{1 - E_0(L1,1) - E_1(L1,1)}\left[\frac{m_{A11}^o m_{A12|H1}^o - m_{A11}^o E_0(L1,2)}{1 - E_0(L1,2) - E_1(L1,2)} - \nu E_0(L1,1)\right] \\ \beta_1^3 + \left(\alpha_1^3 - \beta_1^3\right) p_1 = \frac{1}{\left[1 - E_0(L1,1) - E_1(L1,1)\right]\left[1 - E_0(L1,2) - E_1(L1,2)\right]}\left[\frac{m_{A12}^o m_{A13|HH1}^o - m_{A12}^o E_0(L1,3)}{1 - E_0(L1,3) - E_1(L1,3)} - C\right] \end{cases},$$

which is identified by the same argument as the one used to establish identification in the absence of classification error. Note that the right sides of the expressions in this system consist of known objects or objects already identified. Therefore, three periods of observations on performance allow me to recover α_1, β_1, and p_1.

In the presence of multiple skill types, as before an analogous argument holds with p_1 in the above expressions replaced by the mean initial prior over the skill types of managers who are retained at Level 1 for at least three periods.

2.1.2 Identification of the Output and Human Capital Process

Consider now the problem of identification of $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$, $\Pr(h_{t+1}|h_t, k_t)$, and $\Pr(p_1)$, and of the parameters governing $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$ and $\Pr(h_{t+1}|h_t, k_t)$. I will show here their identification in two steps. In the first step, I will argue that the distribution of heterogeneity can be separately identified from the state transitions and choice components based on the results of Kasahara and Shimotsu (2009) and Hu and Shum (2012). The arguments of Kasahara and Shimotsu (2009) and Hu and

Shum (2012) for the nonparametric identification of dynamic discrete choice models with unobserved heterogeneity imply that a long enough panel dimension, as in my sample, can be sufficient to ensure the nonparametric identification of the discrete choice component of my model. In the second step, I will argue that reduced forms of economic interest of the parameters of output and human capital, which determine the component probabilities identified in the first step, can be identified by standard arguments based on the restrictions implied by the model.

Before proceeding, note that the role of the productivity shocks determining $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$ is auxiliary to the main focus of the estimation exercise. To see why, note that the employment and job assignment decisions on the part of firms, as well as the acceptance decision on the part of a manager with respect to firms' employment offers, are deterministic conditional on beliefs and the manager's acquired human capital. Hence, productivity shocks simply make employment and job choices stochastic from the point of view of the econometrician, conditional on the current period prior and the sequence of past level assignments and performance. Specifically, these shocks ensure that all observed assignments have non-zero probability under the model. For instance, together with the process for beliefs and the classification error in performance ratings, productivity shocks help the model account for observations on managers with the same characteristics (age, education, and year of entry) and history of level assignments and recorded performance ratings, who are assigned at some point to different jobs.

Lastly, note that given $\{\alpha_k, \beta_k\}_{k=1}^3$, $\Pr(p_{t+1}|p_t, k_t)$ is known up to the initial prior distribution, since the law of motion for p_t (Bayesian updating) and its parameters ($\{\alpha_k, \beta_k\}_{k=1}^3$) are known. Thus, the problem of identifying $\Pr(p_{t+1}|p_t, k_t)$ reduces to the problem of identifying the distribution of initial priors, $\Pr(p_1)$. I now turn to examine each step of the argument for the identification of $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$, $\Pr(h_{t+1}|h_t, k_t)$, and $\Pr(p_1)$, and of the parameters governing $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$ and $\Pr(h_{t+1}|h_t, k_t)$.

First Step: Assignment and Human Capital Process and the Initial Prior This first step of the argument for identification relies on the results of Hu and Shum (2012). Hu and Shum consider a general class of dynamic discrete choice models with serially correlated and time-varying unobserved state variables, and prove that conditional choice probabilities, the law of motion for the state, and the distribution of initial conditions are nonparametrically identified. In particular, their result applies to frameworks like mine in which the unobserved state variable (in my case, the prior) is time-varying and can evolve depending on past values of the observed state and choice variables. Hu and Shum (2012) generalize the framework of Kasahara and Shimotsu (2009) by allowing the permanent unobserved heterogeneity component to be updated according to a Markov process.

I will first show that based on their results, $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$, $\Pr(h_{t+1}|h_t, k_t)$, and $\Pr(p_{t+1}|p_t, k_t)$ are identified. I will then discuss the applicability of the identification result of Kasahara and Shimotsu (2009) to my case.

An Argument Based on Hu and Shum (2012). Here I first show how my problem can be cast into the framework of Hu and Shum (2012). Next, I discuss the identification of the processes for the choice variable, $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$, for the observed state variable, $\Pr(h_{t+1}|h_t, k_t)$, and for the unobserved state variable, $\Pr(p_{t+1}|p_t, k_t)$.

To see how my identification problem can be mapped into the framework of Hu and Shum (2012), I borrow their notation and let $W_t = (Y_t, M_t)$ denote the vector of observable variables consisting of the choice variable in period t, Y_t, and of the observed state variables in period t, M_t. Let X_t^* denote the unobserved state variable. Hu and Shum (2012) consider the problem of nonparametric identification of $\Pr(W_t, X_t^*|W_{t-1}, X_{t-1}^*)$ in the special case in which

$$\Pr(W_t, X_t^*|W_{t-1}, X_{t-1}^*) = \Pr(Y_t, M_t, X_t^*|Y_{t-1}, M_{t-1}, X_{t-1}^*) = \Pr(Y_t|Y_{t-1}, M_{t-1}, M_t, X_t^*)$$

$$\cdot \Pr(M_t|Y_{t-1}, M_{t-1}, X_t^*) \Pr(X_t^*|Y_{t-1}, M_{t-1}, X_{t-1}^*),$$

that is, when $\Pr(Y_t|\cdot)$ and $\Pr(M_t|\cdot)$ do not depend on X_{t-1}^*. To see how my problem is an instance of theirs, let $(Y_t, M_t, X_t^*) = (k_t, h_t, p_t) = (k_t, (t-1, k_{t-1}), p_t)$, where t denotes tenure at my firm. So, from (42) it follows that

$$\begin{aligned}\Pr(Y_t, M_t, X_t^*|Y_{t-1}, M_{t-1}, X_{t-1}^*) &= \Pr(k_t, h_t, p_t|k_{t-1}, h_{t-1}, p_{t-1}) \\ &= \Pr(k_t|h_t, p_t) \Pr(h_t|k_{t-1}, h_{t-1}) \Pr(p_t|k_{t-1}, p_{t-1}) \\ &= \Pr(Y_t|M_t, X_t^*) \Pr(M_t|Y_{t-1}, M_{t-1}) \Pr(X_t^*|Y_{t-1}, X_{t-1}^*). \end{aligned} \tag{61}$$

Consider first the case in which the distribution of performance signals, governed by $\{\alpha_k, \beta_k\}_{k=1}^3$, is known, for instance based on the above identification argument. Thus, the problem of identifying $\Pr(p_{t+1}|p_t, k_t)$ reduces to the problem of identifying the distribution of initial priors, $\Pr(p_1)$. For this, observe that the panel dimension of my sample ($T \geq 5$) implies that the identification result of Hu and Shum (2012) applies, which ensures the nonparametric identification of $\Pr(p_1)$. (Suppose the distribution of performance signals is not known. Since Bayesian updating provides the functional form of the state dependence of the process for p_t, the result of Hu and Shum (2012) can also be invoked to establish the nonparametric identification of the distribution of true performance and of the initial priors.)

Lastly, the identification result of Hu and Shum (2012) also implies that the conditional choice probabilities, $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$, as well as the process for the observed state, $\Pr(h_{t+1}|h_t, k_t)$, are nonparametrically identified.

An Argument Based on Kasahara and Shimotsu (2009). An alternative approach to identification follows from Kasahara and Shimotsu (2009). These authors analyze the nonparametric identification of the number of type components and component probabilities of finite mixture dynamic discrete choice models. Their argument covers the case in which choice probabilities are nonstationary and that in which choice probabilities are first-order state-dependent. Their results do not apply when choice probabilities are simultaneously nonstationary and state-dependent. Nonetheless, given the infinite horizon formulation of my model, the class of problems they consider nests mine.

Kasahara and Shimotsu consider models in which the unobserved state is time-invariant. Hence, their results may seem not to apply to frameworks like mine in which the unobserved state (here, the prior) evolves over time. Suppose, however, that the distribution of true performance is known, as implied by the above argument for the identification of the learning process. Bayesian updating, as argued, implies that the state-dependent process for beliefs is known up to the initial prior and the distribution of true performance. Based on these observations, the identification of the discrete choice component of my model reduces to the identification of the distribution of initial priors, that is, the unobserved distribution of types, and of the employment and job assignment probabilities conditional on the type-specific initial priors, that is, the type-specific components. Note that the process $\Pr(h_{t+1}|h_t, k_t)$ for the observed state is directly identified from the data by standard arguments. See, for instance, Rust (1987). Hence, by applying the argument in Kasahara and Shimotsu (2009) with $\{\alpha_k, \beta_k\}_{k=1}^{3}$ known, I can conclude that the distribution of the prior, $\Pr(p_1)$, including the number of support points and their probabilities, and the choice probabilities, $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$, are nonparametrically identified.

Kasahara and Shimotsu (2009) also provide guidance as to when a certain completeness condition for identification is satisfied. Intuitively, when the panel length of the sample is greater than three, this condition amounts to requiring that observed covariates vary sufficiently over time in a way that changes in the covariates induce heterogeneous changes in choice probabilities across types (see Remark 2(i) after Corollary 1 in their paper). In essence, time-varying covariates help the identification of unobserved heterogeneity. In my framework, in which the discrete choice component of the model admits no covariates, the time-series variation in observed choices substitutes for the required time variation of covariates. Specifically, sufficient conditions for identification are: (a) the panel dimension of the sample is greater than twice the number of types minus one, (b) choice probabilities differ across types, and (c) the probability of the first-period choice is strictly positive and different across types. (See Remark 3 at p. 149 in their paper.) It can be shown that my model satisfies all three of these conditions (counting tenure from the second period on). □

The analysis of Hu and Shum (2012) and Kasahara and Shimotsu (2009) neither requires nor exploits the structural interpretation of the components of the mixture model.[4] I turn now to show how in the second step of my argument for identification, these restrictions can be used to identify reduced forms of interest of the model primitives governing the process of output and human capital acquisition.

Second Step: Output and Human Capital Parameters Note first that the assumption that productivity shocks are type I extreme value distributed implies that the probability of assignment to job $k = 1, 2, 3$ at state (p_{it}, h_t), where t denotes tenure and $(t-1, k_{t-1})$ is a sufficient statistic for h_t, can be expressed as

$$\Pr(k_t = k|p_{it}, t-1, k_{t-1}) = \frac{\eta_{k_{t-1}t-1} \exp\{v(p_{it}, t-1, k_{t-1}, k)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v(p_{it}, t-1, k_{t-1}, k')\}} \tag{62}$$

[4] Kasahara and Shimotsu (2009) and Hu and Shum (2012) establish identification under high-level assumptions, and neither exploits the restrictions on outcomes implied by the underlying economic model. In particular, when the unobserved state variable is continuous, the nonparametric identification result by Hu and Shum (2012) relies on higher-level assumptions, like the invertibility assumption and distinctive eigenvalues assumption, which are difficult to verify explicitly for a specific model. (See the discussion in the Appendix of Hu and Shum (2012).)

for $1 \leq k \leq 3$ and

$$\Pr(k_t = 0|p_{it}, t-1, k_{t-1}) = \frac{\eta_{k_{t-1}t-1} \exp\{v(p_{it}, t-1, k_{t-1}, 0)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v(p_{it}, t-1, k_{t-1}, k')\}} + 1 - \eta_{k_{t-1}t-1}$$

since, as discussed in the Appendix in the paper, the match surplus value of firm A satisfies

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

here $k = 0$ corresponds to the reference alternative of separation between the firm and a manager. As discussed below in Section 4, for convenience, I compute the match surplus value first by solving for the match surplus value from the eighth tenure year of a manager on and, then, by solving by backward induction for the match surplus value in the remaining first seven tenure years, using the computed value from the eight tenure year on as the terminal value. I first argue identification in the context of the example of Section 1 and then in the general case.

Example. Recall the example presented in Section 1 and in the paper. The identification of the output and human capital processes at firm A can be established based on Corollary 3 of Magnac and Thesmar (2002). Assume δ is known—I fix it at 0.95 in estimation. Magnac and Thesmar consider a class of dynamic discrete choice models involving the choice among $1, \ldots, K$ alternatives. Under standard assumptions, they show that the following difference between the expected values of two sequences of choices is nonparametrically identified: first, choose alternative $k < K$ now, alternative K tomorrow, and behave optimally thereafter; second, choose K now and tomorrow, and behave optimally thereafter. See Corollary 3 of Magnac and Thesmar (2002) for a formal statement of this result.

To see how their result applies to my framework, consider the case in which firms of type B have more than one job. Let now $V_t^A(p_t, B)$ denote the match surplus value of firm A in period t when the worker is employed by a firm of type B. Then, the match surplus value of firm A in period t can be thought of as the value of the choice among three alternatives: employing the worker at job $A1$, at job $A2$, or having the worker employed by a firm of type B at its offered job in period t, denoted by $k^*_{Bt} = k^*_{Bt}(p_t)$. That is, to apply the result of Magnac and Thesmar, I interpret their alternative K as the alternative of firm A in which the worker is employed by a firm of type B. If so, then Magnac and Thesmar' result implies

$$(1-\delta)y_A(p_1, k) + \delta EV_2^A(p_2, B|p_1, k_{Ak}) - (1-\delta)y_B(p_1, k^*_{B1}) - \delta EV_2^A(p_2, B|p_1, k^*_{B1}) \tag{63}$$

is nonparametrically identified. (In (63), since human capital depends just on tenure and past level assignment, the dependence on h_1 is subsumed in the tenure index.) To see the implications of this result for the identification of the parameters of the output and human capital processes at firm A, assume that $V_t^A(p_t, B)$ can be well approximated by, say, a second-degree polynomial in p_t, for instance,

$$V_t^A(p_t, B) = \sum_{r=0}^{2} [1 + (1-\delta) \cdot (t-1)] v_r p_t^r.$$

As explained in the Appendix in the paper, I follow this approach in estimation since I do not have direct information on the competitors of the firm in my data. Using the expression for $V_2^A(p_2, B)$, the law of motion for beliefs, and the fact that the expected value of the posterior is the prior, $EV_2^A(p_2, B|p_1, k_{Ak})$ equals

$$(2-\delta)v_0+(2-\delta)v_1p_1+(2-\delta)v_2\left\{1+\frac{(\alpha_{Ak}-\beta_{Ak})^2}{[\beta_{Ak}+(\alpha_{Ak}-\beta_{Ak})p_1]\,[1-\beta_{Ak}-(\alpha_{Ak}-\beta_{Ak})p_1]}\,(1-p_1)^2\right\}p_1^2.$$

The expression for $EV_2^A(p_2, B|p_1, k^*_{B1})$ is analogous, with the only difference that the analogue of the term in braces in $EV_2^A(p_2, B|p_1, k_{Ak})$ is now $\{1 + v_2^B(p_1)\,(1-p_1)^2\}$ with job k^*_{B1} rather than job Ak appearing in the relevant expressions, that is, $EV_2^A(p_2, B|p_1, k^*_{B1})$ equals

$$(2-\delta)v_0+(2-\delta)v_1p_1$$

$$+(2-\delta)v_2\left\{1+\frac{(\alpha_{Bk^*_{B1}}-\beta_{Bk^*_{B1}})^2}{[\beta_{Bk^*_{B1}}+(\alpha_{Bk^*_{B1}}-\beta_{Bk^*_{B1}})p_1][1-\beta_{Bk^*_{B1}}-(\alpha_{Bk^*_{B1}}-\beta_{Bk^*_{B1}})p_1]}\,(1-p_1)^2\right\}p_1^2.$$

Suppose that $v_2^B(p_1)$ is well approximated by a polynomial with normalized term of degree zero. Then, it follows that (63) reduces to

$$(1-\delta)(b_{Ak}-b_{k^*_{B1}})+(1-\delta)(c_{Ak}-c_{k^*_{B1}})p_1$$

$$+\delta(2-\delta)v_2\left\{\frac{(\alpha_{Ak}-\beta_{Ak})^2}{[\beta_{Ak}+(\alpha_{Ak}-\beta_{Ak})p_1]\,[1-\beta_{Ak}-(\alpha_{Ak}-\beta_{Ak})p_1]}-v_2^B(p_1)\right\}p_1^2\,(1-p_1)^2\,,$$

which is a polynomial function of p_1. Since δ is known, if p_1 is identified, as argued above, then it is immediate to see that $b_{Ak} - b_{k^*_{B1}}$ and $c_{Ak} - c_{k^*_{B1}}$ are also identified. Moreover, with δ known and p_1 and $\{\alpha_{Ak}, \beta_{Ak}\}_{k=1}^2$ identified, it follows that $v_2^B(p_1)$ is identified too.

Suppose that human capital acquisition takes the form $b_{fkt} = b_{fk} + h_{fk}(k_{t-1})\cdot(t-1)$ and $c_{fkt} = c_{fk} + h_{fk}(k_{t-1})\cdot(t-1)$. In the second period, since the match surplus value problem of firm A is static, it is immediate that the empirical job assignment frequencies identify $y_A(p_2, k) - y_B(p_2, k^*_{B1})$, where

$$y_A(p_2,k)-y_B(p_2,k^*_{B1})=b_{Ak}+h_{Ak}(k_{t-1})-b_{k^*_{B2}}+[c_{Ak}+h_{Ak}(k_{t-1})-c_{k^*_{B2}}]p_2.$$

Thus, the parameters $b_{Ak} + h_{Ak}(k_{t-1}) - b_{k^*_{B2}}$ and $c_{Ak} + h_{Ak}(k_{t-1}) - c_{k^*_{B2}}$ are also identified, for any job k_{t-1} assigned in period 1. Observe that $b_{k^*_{Bt}}$ and $c_{k^*_{Bt}}$ cannot be separately identified from b_{Akt} and c_{Akt}, respectively. Hence, in estimation I set the coefficients on terms of degree zero and one of the polynomial for $V_t^A(p_1, B)$ at zero. I then interpret the estimated b_{Akt} and c_{Akt} as differences between the corresponding parameters of firm A and its competitors at time t. Hence, whenever any such parameter is found to be not significantly different from zero, it follows that such a parameter is the same across firm A and the second-best firm.

Note that an analogous argument would apply if $V_t^A(p_t, B)$ was approximated by a polynomial of degree higher than two with $f(\delta,t)v_r$ denoting the term of degree r, as long as $f(\delta,t)$ is a known function of δ and t. The infinite horizon version of this example works analogously, either interpreted as a finite horizon problem with a long enough horizon or as a truly infinite horizon problem in which the polynomial approximation for $V_t^A(p_t, B)$ can be interpreted as a semiparametric estimator of an unknown, rather than a known, smooth function of p_t. Hence, even in the truly infinite horizon case, standard results ensure identification and that the estimator of $V_t^A(p_t, B)$ has the usual properties. (See Chen (2007).)

General Case. Consider then the match surplus value problem in the first seven years of tenure of a manager, taking as given the match surplus value function from the eighth year of tenure on. First, note that the law of motion for beliefs, human capital, and productivity shocks satisfies the conditional independence assumption common in models of dynamic discrete choice. Specifically, the distribution of future beliefs, human capital, and productivity shocks is independent over time conditional on their current period values. Second, observe that the discount factor and the distribution of the (additive) productivity shocks are known. (In estimation I fixed $\delta = 0.95$.) Finally, recall, as discussed in the paper, that I treat a manager's employment at the second-best competitor of my firm, corresponding to the option of separation between my firm and a manager, as the reference alternative. Hence, the result in Corollary 3 of Magnac and Thesmar (2002) on the nonparametric identification of models of dynamic discrete choice ensures that the objects

$$(1-\delta)\left[b_{kt}(k_{t-1}) + c_{kt}p_{it}\right] + \delta\eta_{kt}E\left[v(p_{it+1}, t, k, 0)|p_{it}, t-1, k_{t-1}, k\right]$$
$$-\delta E\left[v(p_{it+1}, t, k, 0)|p_{it}, t-1, k_{t-1}, 0\right] \tag{64}$$

for $k = 1, 2, 3$ and $1 \leq t \leq 7$ are nonparametrically identified. Recall that in (64) the parameters $b_{kt}(k_{t-1})$ and c_{kt} are interpreted as differences between the corresponding parameters of my firm and the second-best firm.

Recall also from the paper that I specified the match surplus value when $k = 0$ as a polynomial, so $v(p_{it}, t-1, k_{t-1}, 0) = \sum_{v=2}^{V} \nu_{vt}(k_{t-1})p_{it}^v$, since the terms of degrees zero and one are, respectively, subsumed in $b_{kt}(k_{t-1})$ and c_{kt}. As a consequence, (64) can also be rewritten as

$$(1-\delta)\left[b_{kt}(k_{t-1}) + c_{kt}p_{it}\right] + \delta E\left[\eta_{kt}\nu_{2t}(k_{t-1})p_{it+1}^2 + \ldots + \eta_{kt}\nu_{Vt}(k_{t-1})p_{it+1}^V|p_{it}, t-1, k_{t-1}, k\right]$$
$$-\delta\nu_{2t}(k_{t-1})p_{it}^2 - \ldots - \delta\nu_{Vt}(k_{t-1})p_{it}^V. \tag{65}$$

Given that $\Pr(p_1)$ is identified by the argument in the first step and the parameters $\{\alpha_k, \beta_k\}_{k=1}^3$ are identified as proved above, the process for p_{it} conditional on $k = 1, 2, 3$ is also identified. Therefore, with δ known, what is left to show is that the parameters $b_{kt}(k_{t-1})$, c_{kt}, η_{kt}, and $\nu_{vt}(k_{t-1})$, $2 \leq v \leq V$, are identified.

To see this, note that (65) is a polynomial: since choice probabilities are nonparametrically identified, its degree and all its coefficients are identified. In particular, the coefficients of degree zero and one in (65), the parameters $b_{kt}(k_{t-1})$ and c_{kt}, are straightforwardly identified based on (62) and (65) by the proportions of managers who are assigned to Levels 1, 2, and 3, and the hazard rates of managers' transitions across these levels between tenure $t = 1$ and $t = 7$, as in standard (static) discrete choice models.[5]

Consider now the eighth year of tenure. (See the discussion in the Appendix of the paper for details about model specification.) From this period on, as mentioned, the firm solves an infinite-horizon match surplus maximization problem in which, however, only the parameter c_{38} of static expected output is unknown. The parameter c_{38} can easily be recovered from the empirical frequency of separations in $t = 8$ of managers assigned to Level 3 in $t = 7$. Similarly, the exogenous separation rate parameters η_{kt} are identified by the tenure profile of the hazard rate of separation at each level.[6]

2.2 An Example: Local Identification of the Static Discrete Choice Model

I present here a simple example illustrating how the combination of assumptions and implied restrictions of the model provide a source of identification of the mixture discrete choice component of my model. Specifically, I determine here conditions under which a mixture model of static discrete choice with type I extreme value components and fixed (two, for simplicity) number of components is locally identified. The identification of the more general case with multiple components follows the same logic.

Formally, I assume that there exist two unobserved types of individuals $i = 1, 2$ with utilities u_1 and u_2, and denote by $q = \Pr(I = 1)$ the probability that an individual is of type 1. For the following argument not to be trivial, I assume that $q \in (0, 1)$. Denote by $z \in Z \subseteq \mathbb{R}$ an observed individual characteristic (in my case z amounts to age or experience at entry into the firm: it is just sufficient to treat age or experience at entry as a continuous variable for this argument to hold) and by $\rho(\cdot)$ a differentiable function of z. Let $y \in \{0, 1\}$ be the observed discrete choice, which relates to the latent variable y^* as follows

$$y(z) = \begin{cases} 1, \text{ if } y^*(z) = \rho(z) + u_1 I(i = 1) + u_2 I(i = 2) + \epsilon_1 \geq 0 + \epsilon_0 \\ 0, \text{ if } y^*(z) = 0 + \epsilon_0 \geq \rho(z) + u_1 I(i = 1) + u_2 I(i = 2) + \epsilon_1 \end{cases}$$

where ϵ_1 and ϵ_2 are identically and independently distributed type I extreme value disturbances. Denote their cumulative distribution function by G and their probability density function by g. This model implies

$$P(z) = \Pr(y(z) = 1) = q \Pr(\rho(z) + u_1 + \epsilon_1 - \epsilon_0 \geq 0) + (1 - q) \Pr(\rho(z) + u_2 + \epsilon_1 - \epsilon_0 \geq 0)$$

$$= \frac{q}{1 + \exp\{-\rho(z) - u_1\}} + \frac{1 - q}{1 + \exp\{-\rho(z) - u_2\}} = qG(\rho(z) + u_1) + (1 - q)G(\rho(z) + u_2). \tag{66}$$

[5] In practice, as mentioned in the paper, the parameters $\nu_{vt}(k_{t-1})$, $v \geq 2$, which, as it can be seen from (65), are separately identified from the parameters η_{kt}, proved not to be different from zero. In light of this result that $\nu_{vt}(k_{t-1})$, $v \geq 2$, proved insignificantly different from zero, by interpreting my setting as a case in which the continuation value function of the reference alternative is given, by Magnac and Thesmar (2002) it also follows that the static differences between the one-period expected output at my firm and at the second-best firm, net of the productivity shocks, are nonparametrically identified.

[6] Note that in my framework exogenous separations differ from endogenous separations in that the incidence of exogenous separations does not vary with beliefs about a manager's ability or a manager's performance. Indeed, in the data the fraction of separations in each year at each level and tenure is very weakly or practically unrelated to performance or wages. See also the discussion in Baker, Gibbs, and Holmström (1994b, p. 931) on this. This feature of the data suggests that exogenous separations are a quantitatively important determinant of turnover, as estimation results confirm.

Observe that with $G(x) = 1/(1 + \exp\{-x\})$

$$g(x) = \frac{\exp\{-x\}}{(1 + \exp\{-x\})^2}$$

and

$$g'(x) = \frac{-\exp\{-x\}\,(1 + \exp\{-x\})^2 + 2\exp\{-x\}(1 + \exp\{-x\})\exp\{-x\}}{(1 + \exp\{-x\})^4} = \frac{\exp\{-x\}(\exp\{-x\} - 1)}{(1 + \exp\{-x\})^3}.$$

Note for later that

$$\frac{g'(x)}{g(x)} = \frac{\exp\{-x\}(\exp\{-x\} - 1)}{(1 + \exp\{-x\})^3} \cdot \frac{(1 + \exp\{-x\})^2}{\exp\{-x\}} = \frac{\exp\{-x\} - 1}{1 + \exp\{-x\}}.$$

Proposition 4. *Assume that $q \in (0, 1)$ and there exists an open set $Z^* \subseteq Z$ such that for $z \in Z^*$, $\rho'(z) \neq 0$. Then, the parameters $\vartheta = (q, u_1, u_2)$ are locally identified.*

Proof. The proof draws on the well-known equivalence of local identification with positive definiteness of the information matrix. Therefore, in the following I will show the positive definiteness of the information matrix for model (66). The argument builds on Meijer and Ypma (2008) and Fu (2011). I will break the argument in two distinct claims.

Claim 1. *The information matrix $\Upsilon(\vartheta)$ is positive definite if, and only if, there exists no $w \neq 0$ such that $w'\partial P(z)/\partial\vartheta = 0$ for all z.*

Proof. Note that the loglikelihood of an observation (y, z) is

$$L(\vartheta) = y\ln[P(z)] + (1 - y)\ln[1 - P(z)]$$

and the score function is given by

$$\frac{\partial L(\vartheta)}{\partial\vartheta} = y\frac{\partial P(z)/\partial\vartheta}{P(z)} - (1 - y)\frac{\partial P(z)/\partial\vartheta}{1 - P(z)} = \left[\frac{y}{P(z)} - \frac{1 - y}{1 - P(z)}\right]\frac{\partial P(z)}{\partial\vartheta} = \frac{y - P(z)}{P(z)\,[1 - P(z)]} \cdot \frac{\partial P(z)}{\partial\vartheta}.$$

Hence, the information matrix $\Upsilon(\vartheta)$ is given by

$$\Upsilon(\vartheta) = E\left[\frac{\partial L(\vartheta)}{\partial\vartheta} \cdot \frac{\partial L(\vartheta)}{\partial\vartheta'}|z\right] = E\left\{\frac{[y - P(z)]^2}{P(z)^2\,[1 - P(z)]^2} \cdot \frac{\partial P(z)}{\partial\vartheta} \cdot \frac{\partial P(z)}{\partial\vartheta'}|z\right\} = \frac{E[y - P(z)|z]^2}{P^2(z)\,[1 - P(z)]^2}$$

$$\cdot\frac{\partial P(z)}{\partial\vartheta} \cdot \frac{\partial P(z)}{\partial\vartheta'} = \frac{P(z)\,[1 - P(z)]}{P^2(z)\,[1 - P(z)]^2} \cdot \frac{\partial P(z)}{\partial\vartheta} \cdot \frac{\partial P(z)}{\partial\vartheta'} = \frac{1}{P(z)\,[1 - P(z)]} \cdot \frac{\partial P(z)}{\partial\vartheta} \cdot \frac{\partial P(z)}{\partial\vartheta'}.$$

Since $P(z) \in (0, 1)$, if follows that the desired result holds. □

Claim 2. *If $w'\partial P(z)/\partial\vartheta = 0$ for all z, then $w = 0$.*

Proof. Observe that $\partial P(z)/\partial\vartheta$ is given by

$$\begin{cases} \frac{\partial P(z)}{\partial q} = G(\rho(z)+u_1) - G(\rho(z)+u_2) = 0 \\ \frac{\partial P(z)}{\partial u_1} = qg(\rho(z)+u_1) = 0 \\ \frac{\partial P(z)}{\partial u_2} = (1-q)g(\rho(z)+u_2) = 0 \end{cases}.$$

Suppose that $w'\partial P(z)/\partial\vartheta = 0$ for all z for some $w = (w_1, w_2, w_3)$, that is,

$$w_1[G(\rho(z)+u_1) - G(\rho(z)+u_2)] + w_2 qg(\rho(z)+u_1) + w_3(1-q)g(\rho(z)+u_2) = 0.$$

The derivative of this expression with respect to z evaluated at some $z \in Z^*$ is given by

$$w_1[g(\rho(z)+u_1) - g(\rho(z)+u_2)]\rho'(z) + w_2 qg'(\rho(z)+u_1)\rho'(z)$$

$$+w_3(1-q)g'(\rho(z)+u_2)\rho'(z) = 0. \tag{67}$$

Let $r(z) = g(\rho(z)+u_1)/g(\rho(z)+u_2)$. By dividing the left side and the right side of (67) by $g(\rho(z)+u_2)$ and $\rho'(z)$, I obtain

$$w_1\left[\frac{g(\rho(z)+u_1)}{g(\rho(z)+u_2)} - 1\right] + w_2 q\frac{g'(\rho(z)+u_1)}{g(\rho(z)+u_2)} + w_3(1-q)\frac{g'(\rho(z)+u_2)}{g(\rho(z)+u_2)} = 0,$$

from which it follows

$$w_1[r(z)-1] + w_2 q\frac{g'(\rho(z)+u_1)}{g(\rho(z)+u_1)}r(z) + w_3(1-q)\frac{g'(\rho(z)+u_2)}{g(\rho(z)+u_2)} = 0$$

or, equivalently, using the fact that $g'(x)/g(x) = (\exp\{-x\}-1)G(x)$,

$$w_1[r(z)-1] + w_2 q(\exp\{-\rho(z)-u_1\}-1)G(\rho(z)+u_1)r(z)$$

$$+w_3(1-q)(\exp\{-\rho(z)-u_2\}-1)G(\rho(z)+u_2) = 0,$$

which can be rewritten as

$$\underbrace{[w_1 - w_2 q(1-\exp\{-\rho(z)-u_1\})G(\rho(z)+u_1)]}_{A}r(z)$$

$$-\underbrace{[w_1 + w_3(1-q)(1-\exp\{-\rho(z)-u_2\})G(\rho(z)+u_2)]}_{B} = 0. \tag{68}$$

Since $r(z)$ is a non-constant exponential function of z, (68) holds for all $z \in Z^*$ only if both terms A and B in (68) are zero for each $z \in Z^*$, that is, if

$$w_1 - w_2 q\frac{1-\exp\{-\rho(z)-u_1\}}{1+\exp\{-\rho(z)-u_1\}} = 0 \tag{69}$$

and

$$w_1 + w_3(1-q)\frac{1-\exp\{-\rho(z)-u_2\}}{1+\exp\{-\rho(z)-u_2\}} = 0. \tag{70}$$

Now, a necessary condition for (69) and (70) to be zero at all $z \in Z^*$ is that their derivative with respect to z evaluated at any $z \in Z^*$ is zero. Taking the derivative of (69) with respect to z, evaluated at $z \in Z^*$, it follows

$$w_2 q \frac{\exp\{-\rho(z)-u_1\}\rho'(z)(1+\exp\{-\rho(z)-u_1\}) + (1-\exp\{-\rho(z)-u_1\})\exp\{-\rho(z)-u_1\}\rho'(z)}{(1+\exp\{-\rho(z)-u_1\})^2} = 0,$$

which, since $\rho'(z)$ is different from zero by assumption and $1+\exp\{-\rho(z)-u_1\}$ is also different from zero, can be simplified to

$$w_2 q\left(\exp\{-\rho(z)-u_1\} + \exp\{-\rho(z)-u_1\}\exp\{-\rho(z)-u_1\}\right)$$

$$+w_2 q\left(\exp\{-\rho(z)-u_1\} - \exp\{-\rho(z)-u_1\}\exp\{-\rho(z)-u_1\}\right) = 0$$

or, equivalently, $2w_2 q \exp\{-\rho(z)-u_1\} = 0$. Given that $q \in (0,1)$, it follows $w_2 = 0$. Hence, by (69) it also follows that $w_1 = 0$.

Similarly, taking the derivative of (70) with respect to z, evaluated at $z \in Z^*$, it follows

$$w_3(1-q)\frac{\exp\{-\rho(z)-u_2\}\rho'(z)\left(1+\exp\{-\rho(z)-u_2\}\right)}{(1+\exp\{-\rho(z)-u_2\})^2}$$

$$+w_3(1-q)\frac{(1-\exp\{-\rho(z)-u_2\})\exp\{-\rho(z)-u_2\}\rho'(z)}{(1+\exp\{-\rho(z)-u_2\})^2} = 0,$$

which, since $\rho'(z)$ is different from zero by assumption and $1+\exp\{-\rho(z)-u_2\}$ is also different from zero, can be rewritten as

$$w_3(1-q)\left(\exp\{-\rho(z)-u_2\} + \exp\{-2\rho(z)-2u_2\} + \exp\{-\rho(z)-u_2\} - \exp\{-2\rho(z)-2u_2\}\right) = 0$$

or, equivalently, $2w_3(1-q)\exp\{-\rho(z)-u_2\} = 0$. Given that $q \in (0,1)$, it follows $w_3 = 0$. □

In principle this approach could be extended to a dynamic framework, under the assumption that $\partial\rho(z)/\partial\vartheta$ has a strictly positive derivative everywhere in Z^*. Note, however, that this intuition is merely suggestive. One reason is that $\rho(z)$, u_1, and u_2 would typically no longer be linearly separable and would be related to each other through an unknown function, that is, the value function of the problem. These two issues are the core of the well-known nonparametric underidentification of models of dynamic discrete choice.

2.3 Continuous Choice Component of the Model: Wages

Consider now the identification of the wage parameters. Recall from the paper that in my specification of the process for wages, I assume that the coefficients ϖ_{1k}, ϖ_{2k}, and ϖ_{3k} on, respectively, age_n (age at entry of manager n), age_n^2, and edu_n (education at entry of manager n), are equal at Levels 1 and 2, and denote their common value by ϖ_1, ϖ_2, and ϖ_3. I denote by ϖ_{13}, ϖ_{23}, and ϖ_{33} the corresponding coefficients at Level 3. I also restrict the coefficients on the dummies for the year of entry so that $\varpi_{ym} = 0$, $0 \leq m \leq 3$, and $\varpi_{y4} = \varpi_{y5}$.

As for the remaining parameters, recall from the Appendix in the paper that I allow for a tenure effect only at Level 1, parameterized as $\omega_{1t} = \omega_{12} I(t < 5) + \omega_{15} I(t \geq 5)$ 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. Lastly, I assume that the variance of the lognormal error u_{kint} does not vary across skill types at Level 3. (Formally, σ_{ik} is the standard deviation of u_{kint}, which is the sum of the productivity shock at the job offered by a firm with technology f to managers of type i assigned to Level k at my firm in t, ε_{knt}, and of measurement error, ε^m_{kint}.) Hence, the estimated mean wage parameters are $\{\varpi_{0ik}\}_{ik}$, $\{\varpi_j, \varpi_{j3}\}_{j=1}^3$, ω_{12}, $\{\varpi_{ym}\}_{m=5}^9$, and $\{\omega_{2i}\}_{i=1}^4$, whereas the estimated wage variance parameters are $(\{\sigma_{1i}, \sigma_{2i}\}_{i=1}^4, \sigma_3)$. (See the Appendix in the paper.)

To see how these parameters are identified, recall from the expressions in the paper for the estimated wage equation that conditional on beliefs, (log) wages are determined by a linear semiparametric regression model with a random intercept, $\omega_i(h_1, k) + \omega_{1t} \cdot (t-1) I(k = 1)$, which is individual-specific, time-varying, and parameterized by $\{\varpi_{0ik}\}_{ik}$, $\{\varpi_j, \varpi_{j3}\}_{j=1}^3$, ω_{12}, and $\{\varpi_{ym}\}_{m=5}^9$, and with a random slope, $\{\omega_{2i}\}_{i=1}^4$, 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, for instance, Hoderlein, Klemelä, and Mammen (2010). Intuitively, the average (log) wage in the first year of tenure of managers with the same age, education, and year of entry if they entered the firm after 1973, provides information about the parameters $\{\varpi_{0i1}\}_{i=1}^4$ and $\{\omega_{2i}\}_{i=1}^4$. By construction, these parameters flexibly capture the systematic variability of mean log wages at Level 1 in the first year of employment among individuals with the same observed characteristics. As in standard finite mixture models with lognormal components, here $\{\varpi_{0i1}\}_{i=1}^4$ and $\{\omega_{2i}\}_{i=1}^4$ are identified not only by the characteristics of the observed distribution of wages, for instance, the number and location of its modes and its skewness, but also by changes in this level distribution of wages with tenure. Specifically, conditional on $\{p_{i1}\}_{i=1}^4$, $\{\alpha_k, \beta_k\}_{k=1}^3$, and the parameters of classification error, a source of identification for $\{\omega_{2i}\}_{i=1}^4$ is the time variation in the average log wage of managers at a same level of the same skill type (and, thus, initial prior), education, and age, who entered the firm between 1970 and 1973, or in the same year if they entered after 1973, and experience different realized performance leading to different posteriors. The level parameters $\{\varpi_{0i2}, \varpi_{0i3}\}_{i=1}^4$, instead, are identified by the average log wage of managers of the same skill type, age, education, and year of entry if they entered after 1973, who are assigned to Level 2 or Level 3, respectively, compared to individuals with the same characteristics continually assigned to Level 1.

As for the parameters capturing the effect of observable manager characteristics on wages, note that the parameter ω_{12} on the tenure term $t-1$ at Level 1 is identified by the variation with tenure of the average log wage of managers of the same skill type, age, education, and year of entry if they entered after 1973, and with the same history of performance ratings, continually assigned to Level 1. The average log wage of managers entering the firm in the same year, or before 1974, of the same skill type and at a same level but with different age or years of completed education at entry identifies, respectively, ϖ_1, ϖ_2, and ϖ_3, for managers assigned to Levels 1 and 2, and ϖ_{13}, ϖ_{23}, and ϖ_{33}, for managers assigned to Level 3. Similarly, the average log wage of individuals of the same skill type, age, education, and level assignment, who entered the firm between 1974 and 1979, compared to those who entered in earlier years, identifies ϖ_{ym}, $5 \leq m \leq 9$. Lastly, second moments of the distribution of wages at each level pin down $\{\sigma_{1i}, \sigma_{2i}\}_{i=1}^{4}$ at Levels 1 and 2, and σ_3 at Level 3.

3 Data

The original BGH dataset includes 74,071 employee-year observations on 16,133 managers at one U.S. firm over the twenty-year period between 1969 and 1988. BGH report that management constitutes approximately 20 percent of total employment each year. Over the sample years, 12,439 managers enter the firm. (In the sample of 74,071 individuals, 3,694 have missing tenure information when first observed.) The average age of entrants into managerial positions is 33 years, with a standard deviation of approximately 8 years, from a minimum of 20 to a maximum of 71. Their average number of years of education is 15, with a standard deviation of approximately 2 years, from a minimum of 12 to a maximum of 23. Of these 12,439 managers, 3,891 enter the firm between 1970 and 1979, for a total of 30,675 employee-years.

Exit from the firm is substantial in each year. For the sample of entrants into the firm between 1970 and 1979, 10.7 percent leave the firm after one year, 21.1 percent leave after two years, and 60.2 leave by the tenth year. Equivalently, only 39.8 percent of managers have careers lasting 10 years or longer; see Table II in BGH. Overall, only 6,577 managers have missing level information over the sample years, so the total number of observations on individuals at Levels 1–4, 65,851 overall, accounts for 97.6 percent of managers who do not have missing level information, for a total of 67,494 ($= 74,071 - 6,577$) observations. In the original sample, 45,673 individuals have recorded performance ratings, of which 36,750 (80.46 percent) are either 1 or 2.

BGH aggregate job titles into levels according to the pattern and frequency of transitions of managers across titles. Specifically, as explained in detail by BGH, the original data contain 276 different job titles, but 14 titles, each representing at least 0.5 percent of employee-years, comprise about 90 percent of all observations and 93 percent of those with titles coded. In order to fill the job ladder from the bottom to the top of the firm's hierarchy, BGH add to these 14 titles the title of Chairman-CEO and the only two titles observed in transitions from the 14 major titles to the position of Chairman-CEO, leading to a total of 17 titles. Then, BGH construct transition matrices to analyze movements of employees between these 17 titles, both for individual years and over the sample period.

Based on these transitions, BGH construct eight hierarchical levels. According to the procedure that BGH follow, *Level 1* consists of the three titles that employ almost only new hires. Most transitions from Level 1 within the firm are to six other titles, identified as *Level 2*. Transitions out of Level 2 are almost exclusively to three other job titles, classified as *Level 3*. After major titles have been assigned to levels, less common titles have been allocated to levels based on observed movements between them and titles already assigned. This process is continued until all 17 titles are assigned to a level.

The literature on the internal economics of the firm commonly argues that higher-level jobs of a firm's hierarchy correspond more to general management, whereas lower-level jobs depend more on specialized functional knowledge and require performing less complex tasks. This pattern of the task content of jobs at different levels of a firm's hierarchy is present in the BGH data. For instance, as described by BGH, at Levels 1–4 about 60 percent of the jobs relate to specific 'line' (revenue-generating) business units, positions that involve direct contact with customers or creating and selling products. Approximately 35 percent are 'staff' or 'overhead' positions in areas such as *Accounting*, *Finance*, or *Human Resources*. At Levels 5–6 these two percentages decrease to 45 and 25 percent, respectively, whereas general management descriptions such as 'General Administration' or 'Planning' increase to about 30 percent. At Levels 7–8 all jobs are of this latter type, and they entail managing large groups, coordinating across business units, and strategic planning.

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

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

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

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

4 Baseline Estimation

I derive here the likelihood function, provide details about the numerical solution of the model, present Monte Carlo evidence about the identifiability of the model's parameters in practice, and, finally, discuss the implications of the parameter estimates reported in the paper for the informativeness of the jobs of competitors of my firm in the market for managers.

4.1 Likelihood Function

I estimate the vector of model parameters, ϑ, by full-information, full-solution, nonparametric maximum likelihood. The loglikelihood function for the sample is derived as follows.

Formally, let $e_n = (age_n, edu_n, year_n)$ denote the vector of characteristics of manager n at entry into the firm, which consists of the manager's age (age_n), years of completed education (edu_n), and year of entry into the firm ($year_n$). Recall that, in light of the high separation rate in each year and tenure in my data, I restrict attention to the first eight years of tenure of a manager at the firm. Specifically, for each manager I compute the probability of the observed assignment and wage in each tenure up to tenure $t = 8$ (included) and the probability of the observed performance rating up to tenure $t = 7$ (included). Let then $T_n = \min\{\overline{T}_n, 8\}$ be the length of the observation period for manager n, corresponding to the minimum between the last year of tenure of the manager at the firm ($\overline{T}_n$) and the eighth year of tenure. By the same convention adopted by BGH, here the event in which level assignment and performance rating are simultaneously first missing is interpreted as a separation.

Let $O_{nt} = (L^o_{nt}, w^o_{nt}, R^o_{nt})$ denote manager n's outcome in period t and $O_{int} = (L^o_{int}, w^o_{int}, R^o_{int})$ the manager's period t outcome when of type i. Here $L^o_{nt}, L^o_{int} \in \{0, 1, 2, 3\}$ represent the observed level assignment in period t for manager n and for manager n of type i, respectively. Recall that the assignment to Level 0 corresponds to a separation. Similarly, $w^o_{nt}, w^o_{int} \in \{\emptyset\} \cup \{\mathbb{R}_+\}$ denote the observed wage, possibly missing, and $R^o_{nt}, R^o_{int} \in \{\emptyset, 0, 1\}$ the observed performance rating, possibly missing, in period t for manager n and for manager n of type i, respectively. Recall from the paper that R_{nt} denotes the performance realized in period t for manager n (unobserved by the econometrician); R_{int} is similarly defined when the manager is of type i. Thus, the probability of manager n's outcome history $O_n = (O_{n1}, \ldots, O_{nT_n})$ conditional on $e_n = (age_n, edu_n, year_n)$ can be expressed as

$$\Pr(O_{n1}, \ldots, O_{nT_n}|e_n) = \sum_{i=1}^{I} \Pr(i|e_n) \sum_{\theta \in \{\alpha,\beta\}} \Pr(\theta|i, e_n) \Pr(O_{in1}, \ldots, O_{inT_n}|\theta, i, e_n)$$

$$= \sum_{i=1}^{I} q_i \left[p_{in1} \Pr(O_{in1}, \ldots, O_{inT_n}|\alpha, i, e_n) + (1 - p_{in1}) \Pr(O_{in1}, \ldots, O_{inT_n}|\beta, i, e_n)\right], \tag{71}$$

where $I = 4$, as discussed in the paper, $\Pr(i|e_n) = q_i$, $\Pr(\alpha|i, e_n) = p_{in1}$, and $\Pr(\beta|i, e_n) = 1-p_{in1}$. Note that since a manager's ability is unknown to the econometrician, a manager's likelihood contribution is obtained by integrating over the two possible unobserved ability levels of the manager. Similarly, because the prior belief about a manager's ability and a manager's wage depend on the manager's skill type, also unobserved by the econometrician, computing the likelihood contribution of a manager's outcome history requires integration over the manager's possible skill types.

The probability of an observed assignment is computed as follows. First, recall that I maintain that level assignment is measured in the data without error, so the observed level assignment for a manager in any given tenure corresponds to the firm's preferred choice in that period. Second, note that the assumed process for recorded performance ratings implies that, conditional on a manager's true performance, observed performance has no impact on level assignment. The reason is that conditional on true performance, recorded performance is independent of a manager's ability, beliefs about it, or a manager's human capital—except for tenure in the firm, already part of the observed part of the state. Thus, recorded performance does not provide any additional information about a manager's ability (or skill type) besides the information provided by true performance. Third, according to the model, because neither the firm nor a manager observe the manager's ability θ, the firm's assignment policy and the manager's job acceptance policy depends on only the current posterior that a manager is of high ability (which is just a function of the initial prior and the sequence of past level assignments and realized performance), on the accumulated human capital (which is just a function of tenure in the firm and previous period level assignment), and on the current vector of productivity shocks.

Formally, let $p_{int} = \varphi(p_{in1}|L^o_{in1}, R_{in1}, \ldots, L^o_{int-1}, R_{int-1})$ denote the updated or posterior belief in period t that a manager of skill type i is of high ability, from the prior p_{in1} and the history of past level assignments $(L^o_{in1}, \ldots, L^o_{int-1})$ and realized performance $(R_{in1}, \ldots, R_{int-1})$. The above observations then imply that

$$\Pr(L^o_{int}|L^o_{in1}, R^o_{in1}, R_{in1}, \ldots, L^o_{int-1}, R^o_{int-1}, R_{int-1}, \theta, i, e_n)$$
$$= \Pr(L^o_{int}|\varphi(p_{in1}|L^o_{in1}, R_{in1}, \ldots, L^o_{int-1}, R_{int-1}), t-1, L^o_{int-1}), \tag{72}$$

where $\varphi(p_{in1}|L^o_{in1}, R_{in1}, \ldots, L^o_{int-1}, R_{int-1})$ is given by

$$\frac{\alpha_{L^o_{in1}}^{R_{in1}}(1-\alpha_{L^o_{in1}})^{1-R_{in1}} \cdots \alpha_{L^o_{int-1}}^{R_{int-1}}(1-\alpha_{L^o_{int-1}})^{1-R_{int-1}} p_{in1}}{\alpha_{L^o_{in1}}^{R_{in1}}(1-\alpha_{L^o_{in1}})^{1-R_{in1}} \cdots \alpha_{L^o_{int-1}}^{R_{int-1}}(1-\alpha_{L^o_{int-1}})^{1-R_{int-1}} p_{in1} + \beta_{L^o_{in1}}^{R_{in1}}(1-\beta_{L^o_{in1}})^{1-R_{in1}} \cdots (1-p_{in1})}$$

by Bayes' rule, with $\alpha_{L^o_{in\tau}} \in \{\alpha_1, \alpha_2, \alpha_3\}$ and $\beta_{L^o_{in\tau}} \in \{\beta_1, \beta_2, \beta_3\}$, $\tau = 1, \ldots, t-1$. Note that the dependence of p_{int} on the sequence of past level assignments is due to the fact that the distribution of performance is allowed to vary in the job a manager performs—jobs are differentially informative about ability. Also, recall that the parameters $\{\alpha_k, \beta_k\}_{k=1}^3$ governing the output signals about ability are assumed to be independent of a manager's skill type and invariant over time.

As for wages, according to the model a manager's wage in a period only depends on the manager's current level assignment, prior, skill type, observed characteristics at entry into the firm as recorded by e_n, and tenure in the firm. Thus, I denote the probability density function of the observed wage w_{int}^o in period t for manager n of type i assigned to job L_{int}^o by $f(w_{int}^o | L_{int}^o, p_{int}, i, e_n, t)$.

As for performance ratings, recall that realized performance is unobserved by the econometrician. Also, for the econometrician, the joint likelihood of the observed and true performance ratings of a manager in a period only depends on a manager's current assignment, tenure, and ability. (In particular, this likelihood does not depend on the prior about a manager's ability, conditional on the manager's ability.) Therefore, for any $t \in \{1, \ldots, T_n\}$,

$$\Pr(R_{int}^o, R_{int} | L_{in1}^o, R_{in1}^o, R_{in1}, \ldots, L_{int-1}^o, R_{int-1}^o, R_{int-1}, L_{int}^o, \theta, i, e_n)$$

$$= \Pr(R_{int}^o, R_{int} | L_{int}^o, t, \theta) = \Pr(R_{int}^o | R_{int}, L_{int}^o, t) \Pr(R_{int} | L_{int}^o, \theta).$$

Based on these observations, the likelihood of the outcome history $(O_{in1}, \ldots, O_{inT_n})$, conditional on θ, i, and e_n, for manager n of type i is given by

$$\Pr(O_{in1}, \ldots, O_{inT_n} | \theta, i, e_n) = \Pr(L_{in1}^o, w_{in1}^o, R_{in1}^o, \ldots, L_{inT_n}^o, w_{inT_n}^o, R_{inT_n}^o | \theta, i, e_n)$$

$$= \sum_{R_{in1}} \sum_{R_{in2}} \cdots \sum_{R_{inT_n}} \Pr(L_{in1}^o, w_{in1}^o, R_{in1}^o, R_{in1}, \ldots, L_{inT_n}^o, w_{inT_n}^o, R_{inT_n}^o, R_{inT_n} | \theta, i, e_n),$$

which can be also expressed as

$$\Pr(O_{in1}, \ldots, O_{inT_n} | \theta, i, e_n) = \sum_{R_{in1}} \sum_{R_{in2}} \cdots \sum_{R_{inT_n}} \Pr(L_{in1}^o | p_{in1}) f(w_{in1}^o | L_{in1}^o, p_{in1}, i, e_n, 1)$$

$$\cdot \Pr(R_{in1}^o, R_{in1} | L_{in1}^o, 1, \theta) \cdots \Pr(L_{inT_n}^o | \varphi(p_{in1} | L_{in1}^o, R_{in1}, \ldots, L_{inT_n-1}^o, R_{inT_n-1}), T_n - 1, L_{inT_n-1}^o)$$

$$\cdot f(w_{inT_n}^o | L_{inT_n}^o, \varphi(p_{in1} | L_{in1}^o, R_{in1}, \ldots, L_{inT_n-1}^o, R_{inT_n-1}), i, e_n, T_n) \Pr(R_{inT_n}^o, R_{inT_n} | L_{inT_n}^o, T_n, \theta). \quad (73)$$

Finally, the sample likelihood is the product of the probabilities in (71) over the N managers:

$$\mathcal{L}(\vartheta | e_1, \ldots, e_N) = \prod_{n=1}^{N} \sum_{i=1}^{I} \Pr(i | e_n) \sum_{\theta \in \{\alpha, \beta\}} \Pr(\theta | i, e_n) \Pr(O_{in1}, \ldots, O_{inT_n} | \theta, i, e_n).$$

To compute the estimated value of ϑ, I employ a standard nested fixed-point algorithm that relies on the repeated full solution of the employing firm's match surplus value problem at each trial parameter vector. The optimization algorithm I use to maximize the loglikelihood function is a straightforward implementation of the downward simplex method. Finally, I compute asymptotic standard errors based on the outer product of the scores of the loglikelihood function. I performed all numerical routines in FORTRAN90. At the estimated parameter vector, the loglikelihood for the sample is 752.871.[7]

4.2 Numerical Solution of the Model

My numerical approach to computing the match surplus value (of the firm in my data or the planner, depending on the relevant market configuration) and the job-specific match surplus values determining the probabilities of interest relies on the work of Rust (1987, 1988, 1994) on the solution and estimation of stochastic dynamic discrete choice problems. Here I describe how I apply Rust's method to the equilibrium best-response employment and job assignment problem of my firm.

4.2.1 Match Surplus Value Problem

I formulate assumptions that ensure that this problem is stationary from tenure $t = 8$ on. Given these assumptions, I can break the problem into one stationary problem from tenure $t = 8$ on and seven non-stationary problems, one for each of the tenures 1 through 7. Of course, the (expected present discounted) continuation value at tenure 1 is the value at tenure 2, and so on. In particular, the continuation value at tenure 7 is the stationary value at tenure 8.

To ensure the stationarity of the match surplus value problem from tenure 8 on, I make two assumptions. First, I assume that from tenure $t = 8$ on, the human capital acquired by a manager has the same productive value, regardless of a manager's employment history at the firm. (The reason is that, due to the high rate of attrition, the sample contains only a small number of observations on managers at high tenures and the employment outcomes of these managers from $t = 8$ on display little variation. So, the estimation of different human capital parameters from $t = 8$ on for managers with different outcome histories at the firm proved unfeasible.) Thus, $y(p_{it}, t-1, k_{t-1}, k) = y(p_{it}, k)$ at $t \geq 8$. Second, I assume that from tenure 8 on, the separation shocks are independent of tenure at each job, and I denote their common value across these tenures at job k by η_k.

Consider now the stationary match surplus value problem from the eighth year of tenure on. For simplicity, omit the firm and tenure subscripts, and the manager n subscript from the relevant variables. Denote by $\boldsymbol{\varepsilon}$ the current value of productivity shocks and by $\boldsymbol{\varepsilon}'$ their future value. Then, the match surplus value of firm A at $t \geq 8$ is

$$V_8(p_i, \boldsymbol{\varepsilon}) = \max_{k \in \{0,1,2,3\}} V_8(p_i, \boldsymbol{\varepsilon}, k) = \max_{k \in \{0,1,2,3\}} \{v_8(p_i, k) + (1-\delta)\varepsilon_k\}, \tag{74}$$

where, for Level $k \in \{1, 2, 3\}$,

$$v_8(p_i, k) = (1-\delta) y(p_i, k) + \delta \eta_k r_k(p_i) \int_{\varepsilon'} V_8(P_{Hk}(p_i), \boldsymbol{\varepsilon}') dG$$

[7] Note that, given the two-parameter lognormal assumption for the distribution of wages at each level, the actual wage for each manager in each year is a constant that can be factored out in computing the likelihood. I follow this convention in reporting the likelihood value.

$$+\delta\eta_k[1-r_k(p_i)]\int_{\varepsilon'} V_8(P_{Lk}(p_i),\boldsymbol{\varepsilon}')dG, \tag{75}$$

$r_k(p_i) = \alpha_k p_i + \beta_k(1-p_i)$, and the value of separation, $v_8(p_i, 0)$, is approximated by a polynomial as discussed in the paper. For tenures t ranging from 1 through 7, instead, the match surplus value problem of the firm has state $s_{it} = (p_{it}, t-1, k_{t-1})$, as discussed in the paper, and value

$$V(p_{it}, t-1, k_{t-1}, \boldsymbol{\varepsilon}_t) = \max_{k\in\{0,1,2,3\}}\{V(p_{it}, t-1, k_{t-1}, \boldsymbol{\varepsilon}_t, k)\} = \max_{k\in\{0,1,2,3\}}\{v(p_{it}, t-1, k_{t-1}, k) + (1-\delta)\varepsilon_{kt}\}, \tag{76}$$

where

$$v(p_{it}, t-1, k_{t-1}, k) = (1-\delta)y(p_{it}, t-1, k_{t-1}, k) + \delta\eta_{kt} r_k(p_{it})\int_{\boldsymbol{\varepsilon}_{t+1}} V(P_{Hk}(p_{it}), t, k, \boldsymbol{\varepsilon}_{t+1})dG$$

$$+\delta\eta_{kt}[1-r_k(p_{it})]\int_{\boldsymbol{\varepsilon}_{t+1}} V(P_{Lk}(p_{it}), t, k, \boldsymbol{\varepsilon}_{t+1})dG \tag{77}$$

with $V(P_{jk}(p_{it}), t, k, \boldsymbol{\varepsilon}_{t+1}) = V_8(P_{jk}(p_{it}), \boldsymbol{\varepsilon}_{t+1})$, $j = H, L$, when $t = 7$.

4.2.2 Algorithm

I turn now to the numerical calculation of the match surplus value. Under the assumption that the shocks $\boldsymbol{\varepsilon}_t = (\varepsilon_{0t}, \varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})$ have joint conditional (on p_{it}) multivariate type I extreme value distribution, their density is given by

$$f(\varepsilon_{0t}, \varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}|p_{it}) = \Pi_{k=0}^{3}\exp(-\varepsilon_{kt}-\gamma)\exp[-\exp(-\varepsilon_{kt}-\gamma)],$$

where $\gamma = 0.5772$ is the Euler constant. Recall that the density function of a type I extreme value distribution is $f(x) = \frac{1}{b}\exp(\frac{-x+a}{b})\exp[-\exp(\frac{-x+a}{b})]$, with mean $E(x) = a + \gamma b$ and variance $V(x) = b^2\pi^2/6$. For all shocks to have mean zero and variance $\pi^2/6$, as assumed in the paper, the location parameter of the distribution of each shock, a, must equal $-\gamma$ and the variance parameter, b, must equal 1.

Under this distributional assumption, in any tenure the match surplus value problem is akin to a standard dynamic multinomial logit problem. Hence, standard techniques can be applied to derive the probabilities of observed job assignments, including separation. I solve for the probability of an observed assignment in three steps, as follows. In the first step I solve for the match surplus value function at tenure $t \geq 8$. In the second step I use this computed value function as the terminal value in a backward induction recursion that solves for the match surplus value function from tenure $t = 1$ to $t = 7$. In the third step I derive the probabilities of interest.

First, I compute the match surplus value function in tenure $t \geq 8$. Note that at any prior p_i',

$$\int_{\boldsymbol{\varepsilon}'} V_8(p_i', \boldsymbol{\varepsilon}')dG = \int_{\boldsymbol{\varepsilon}'}\max_{k'\in\{0,1,2,3\}}\{v_8(p_i', k') + \varepsilon_{k'}'\}dG = \ln\sum\nolimits_{k'\in\{0,1,2,3\}}\exp\left[v_8(p_i', k')\right], \tag{78}$$

which implies that $v_8(p_i, k)$ from (75) can be rewritten as

$$v_8(p_i, k) = (1-\delta)y(p_i, k) + \delta\eta_k r_k(p_i) \ln \sum\nolimits_{k' \in \{0,1,2,3\}} \exp\left[v_8(P_{Hk}(p_i), k')\right]$$

$$+\delta\eta_k[1 - r_k(p_i)] \ln \sum\nolimits_{k' \in \{0,1,2,3\}} \exp\left[v_8(P_{Lk}(p_i), k')\right]. \tag{79}$$

To complete this step, given the approximation for $v_8(p_i, 0)$, I solve a three-dimensional contraction mapping problem, with contraction

$$\Gamma(v_8)(p_i, k) = (1-\delta)y(p_i, k) + \delta\eta_k r_k(p_i) \ln \sum\nolimits_{k' \in \{0,1,2,3\}} \exp\left[v_8(P_{Hk}(p_i), k')\right]$$

$$+\delta\eta_k[1 - r_k(p_i)] \ln \sum\nolimits_{k' \in \{0,1,2,3\}} \exp\left[v_8(P_{Lk}(p_i), k')\right],$$

$k = 1, 2, 3$, where Γ is an operator on the function v_8. Note that here I follow the formulation in Rust (1988, 1994), in which the functional operator to be solved for is defined as a fixed point of the (expected present discounted) value of choosing an action, rather than the related formulation of Rust (1987), in which that operator is defined as a fixed point of the (expected present discounted) continuation value of choosing an action.

The second step in solving for the probability of an observed assignment relies on the numerical solution of the match surplus value function at $t \geq 8$ as input to the backward induction recursion defining the match surplus value functions in the remaining tenure dates. Specifically, consider tenures between $t = 2$ and $t = 7$. At these tenures, given (76)–(79), I can compute $v(p_{it}, t-1, k_{t-1}, k)$ as

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

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

where the continuation value at $t = 7$ is $v(P_{jk}(p_{it}), t, k, k') = v_8(P_{jk}(p_{it}), k')$, $j = H, L$. Next, consider the match surplus value functions at $t = 1$. These value functions differ from the value functions just derived because the state only consists of p_{i1}.[8]

For the third step, I compute the probability of an observed assignment, including separation, using the match surplus value functions at each tenure calculated as above. By Rust (1994), the probability of the observed assignment $k_t = k$ for a manager of type i in any tenure between $t = 2$ and $t = 7$ is given by

$$\Pr(k_t = k | p_{it}, t-1, k_{t-1}) = \frac{\eta_{k_{t-1}t-1} \exp\{v(p_{it}, t-1, k_{t-1}, k)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v(p_{it}, t-1, k_{t-1}, k')\}} \tag{80}$$

for $1 \leq k \leq 3$, and

$$\Pr(k_t = 0 | p_{it}, t-1, k_{t-1}) = \frac{\eta_{k_{t-1}t-1} \exp\{v(p_{it}, t-1, k_{t-1}, 0)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v(p_{it}, t-1, k_{t-1}, k')\}} + 1 - \eta_{k_{t-1}t-1} \tag{81}$$

[8] Given the varying size of the output level parameters in estimation, continuation values in the relevant functional equations are computed using the fact that $\log(e^{x_1} + e^{x_2}) = \log\left[e^{y-y}(e^{x_1} + e^{x_2})\right] = y + \log(e^{x_1 - y} + e^{x_2 - y})$.

for $k = 0$. The probability of assignment k, $0 \leq k \leq 3$, at $t = 1$ is given by

$$\Pr(k_1 = k|p_{i1}) = \frac{\exp\{v(p_{i1}, k)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v(p_{i1}, k')\}}. \tag{82}$$

The probability of assignment k, $1 \leq k \leq 3$, at $t \geq 8$ is given by

$$\Pr(k_t = k|p_{it}) = \frac{\eta_{k_{t-1}t-1} \exp\{v_8(p_{it}, k)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v_8(p_{it}, k')\}}, \tag{83}$$

and the probability of assignment $k = 0$ at $t \geq 8$ is given by

$$\Pr(k_t = 0|p_{it}) = \frac{\eta_{k_{t-1}t-1} \exp\{v_8(p_{it}, 0)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v_8(p_{it}, k')\}} + 1 - \eta_{k_{t-1}t-1}. \tag{84}$$

I compute the job-specific match surplus value functions $v_8(p_i, k)$ by value function iteration. I discretize the support of p_i, $[0, 1]$, to a uniform grid of 100 equidistant points. (I have also experimented with finer grids, but results are virtually unaffected and the increase in computational cost is substantial.) Observe that in my setup the process for beliefs is much richer than is often assumed in (binary signal) learning models in which all actions (here, jobs) are equally informative and information is symmetric across high and low states of nature (here, managers of high and low ability). These models, in fact, feature only one learning parameter for all jobs, because jobs are equally informative, that is, $\alpha = \alpha_k$ and $\beta = \beta_k$, and, by symmetry, $\alpha = 1 - \beta$. Thus, starting with a prior p_{it} at tenure t, the posterior belief reached after the experience of a success and a failure in a job, or after the experience of a failure and a success, would be $P_L(P_H(p_{it})) = P_H(P_L(p_{it})) = p_{it}$.

I assume neither that all jobs are equally informative about ability nor that information is symmetric across managers of high and low ability to allow for a flexible belief process. I rely on a nearest-neighborhood procedure to ensure that the posterior p_{it+1} that a manager is of high ability, computed for each possible prior p_{it} on the uniform grid for the interval $[0, 1]$, is a point on the same grid.

4.3 Monte Carlo Analysis

The model relies on a multidimensional nonlinear maximization routine to implement the maximum likelihood estimator. I now discuss evidence from a number of simulation-based experiments conducted in order to investigate the practical identifiability of the model's parameters. For these experiments, I simulate 1,426 realizations of the shocks (the size of the estimation sample) 50 times with each parameter in ϑ set equal to its estimated value. Next, I reestimate the model on each simulated dataset. Then, I compare the estimates obtained on these simulated data with the estimates obtained on the actual data. Table A.1 displays statistics on the sample distribution of the parameter estimates across the 50 simulated datasets.

Formally, denote by $\widehat{\vartheta}_{pd}$ the estimated value of the parameter ϑ_p, $1 \leq p \leq 75$, based on dataset $d \in \{1, \ldots, 50\}$, by $\widehat{\vartheta}_p$ its mean estimated value across the 50 datasets, by $\sigma_{\widehat{\vartheta}_p}$ the sample standard deviation of $\widehat{\vartheta}_{pd}$ across the 50 datasets, and by $\widehat{\sigma}_{\widehat{\vartheta}_{pd}}$ the asymptotic standard error of the parameter ϑ_{pd} estimated on the d-th dataset. In the second column of Table A.1, I report the estimate of each parameter based on the original sample of 1,426 individuals and in the third column, the simulation bias, that is, the average deviation of each estimated parameter from its true value (that is, the value estimated based on the original data) across the 50 experiments. Namely, I compute this bias as

$$bias = \widehat{\vartheta}_p - \vartheta_p = \frac{1}{50}\sum\nolimits_{d=1}^{50} \widehat{\vartheta}_{pd} - \vartheta_p.$$

In the fourth column of Table A.1, I report the t-statistic of this bias,

$$t\text{-}statistic\ bias = \sqrt{50}\left(\frac{\widehat{\vartheta}_p - \vartheta_p}{\sigma_{\widehat{\vartheta}_p}}\right),$$

where the average or sample standard deviation $\sigma_{\widehat{\vartheta}_p}$ of the estimated parameter $\widehat{\vartheta}_{pd}$ over the 50 experiments is computed as

$$\sigma_{\widehat{\vartheta}_p} = \sqrt{\frac{1}{49}\sum\nolimits_{d=1}^{50}\left(\widehat{\vartheta}_{pd} - \frac{1}{50}\sum\nolimits_{d=1}^{50}\widehat{\vartheta}_{pd}\right)^2}.$$

I report the values of $\sigma_{\widehat{\vartheta}_p}$ in the fifth column of Table A.1. Finally, in the sixth column of that table I report the average estimated standard error of each parameter estimate, where the standard error $\widehat{\sigma}_{\widehat{\vartheta}_{pd}}$ of the parameter estimate $\widehat{\vartheta}_{pd}$ is obtained from the outer product of the scores of the loglikelihood function for the d-th simulated dataset. I compute this mean estimated standard error as

$$E(\widehat{\sigma}_{\widehat{\vartheta}_{pd}}) = \frac{1}{50}\sum\nolimits_{d=1}^{50}\widehat{\sigma}_{\widehat{\vartheta}_{pd}}.$$

Observe that biases overall seem quite small and mostly precisely estimated. The only parameters for which the bias seems at all significant are c_{12}, c_{23}, c_{25}, c_{26}, $b_{37}(L3) - b_{37}(L2)$, and c_{38}. But for all of these parameters, the bias is negligible as a fraction of the parameter values.

Since the model features several dimensions of heterogeneity and I do not have direct information about a manager's output at the firm, estimating the output parameters governing job assignment choices might be expected to be difficult. Yet, based on the empirical standard deviations of the parameters across the 50 experiments, it is apparent that most of the model's parameters are precisely estimated (these estimated values equal the baseline estimates in the second column of Table A.1 plus the biases in the third column). Standard errors based on the Hessian matrix are only slightly understated, with the exception of the parameters c_{12}, c_{23}, and c_{38}, which are significantly overstated. Overall, I interpret the results of this Monte Carlo exercise as providing evidence in support of the model being identified.

4.4 Implications of Parameter Estimates: Information Bounds

In the paper I have focused on the implications of my estimates for the characteristics of the jobs of the firm in my data. Here I argue that I can derive lower and upper bounds on the informativeness of jobs at my firm's (best) competitor for a manager, as measured by the likelihood ratio of high output between a manager of low and high ability, β_n/α_n, based on the estimates of the parameters of the wage process at my firm. Recall that, by the symmetry assumption, for a given manager n, the probability of success of a manager assigned to a job of a competitor of my firm is α_n, if the manager is of high ability, and β_n, if the manager is of low ability.

To derive these bounds, I exploit the model's implication that the wages paid by my firm are the sum of the one-period expected output of a manager at the best competitor of my firm and of a compensating differential. In turn, the best competitor's expected output embedded in paid wages is informative about the distribution of true performance at its jobs. Recall the expression for paid wages in the paper. For simplicity, let $\psi_{0kint} \equiv \psi_i(h_{n1}, k_{At}) + \psi_{k_{At}} \cdot (t-1)$ and $\psi_{1i} \equiv \psi_{1i}(k_{At})$. For simplicity, let $\omega_{1kt} = \omega_{1t}$ when the assigned job is Level 1 and $\omega_{1kt} = 0$ otherwise. The formal result is contained in the following:

Proposition 5. *If $\psi_{1i} \leq 0$ and $y_{f_{0t}Lk}(h_t) + \psi_{0kint} \geq 0$, then*

$$\frac{\beta_n}{\alpha_n} \leq \frac{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1)}{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) + \omega_{2i}}. \tag{85}$$

Instead, if $\psi_{1i} \geq 0$ and $y_{f_{0t}Lk}(h_t) + \psi_{0kint} \leq 0$, then

$$\frac{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1)}{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) + \omega_{2i}} \leq \frac{\beta_n}{\alpha_n}. \tag{86}$$

Note that if $\alpha_n \geq \beta_n$, then the ratio β_n/α_n ranges from zero to one. In practice, based on the parameter estimates, the ratio β_n/α_n on the left side of (85) and on the right side of (86) ranges between 0.795 and 0.886 for managers assigned to Level 1, between 0.802 and 0.886 for managers assigned to Level 2, and between 0.806 and 0.887 for managers assigned to Level 3. Observe that, by Bayes' rule,

$$P_{f_{0t}Hk}(p_{int}) = \frac{p_{int}}{p_{int} + (1 - p_{int})\beta_n/\alpha_n} \text{ and } P_{f_{0t}Lk}(p_{int}) = \frac{p_{int}}{p_{int} + (1 - p_{int})(1 - \beta_n)/(1 - \alpha_n)}$$

so that updated probabilities after success depend on only the ratio β_n/α_n. Thus, based on (85) and (86), I can compute lower and upper bounds on the number of years that the market would take in order to learn about a manager's ability, if a manager were employed at the best competitor of my firm rather than at my firm. Starting from an average prior of 0.473 across the four manager skill types (that is, $\sum_i q_i p_{i1} = 0.473$ based on the estimates in the paper), I estimate that it would take between 11 and 20 consecutive years of high output at the best competitor of my firm for this prior to converge to 0.90. At my firm, this number ranges between 20 years at Level 1 and 23 years at Level 2 or 3. Hence, analogously to the findings about the speed of learning at my firm discussed in the paper, learning at the best competitor of my firm also occurs slowly, albeit somewhat faster than at my firm.

The proof of Proposition 5 is as follows. Recall that, by definition, firm f_{0t}'s expected output at state s_{int} and job k_{0t}, net of productivity shocks, is given by

$$y_{f_{0t}}(s_{int}, k_{0t}) = y_{f_{0t}Lk}(h_t) + \beta_n \left[y_{f_{0t}Hk}(h_t) - y_{f_{0t}Lk}(h_t)\right] + (\alpha_n - \beta_n)\left[y_{f_{0t}Hk}(h_t) - y_{f_{0t}Lk}(h_t)\right] p_{int}. \quad (87)$$

From the expressions for paid wages in the paper,

$$\ln(w^o_{Aint}) = \omega_i(h_{n1}, k) + \omega_{1t} \cdot (t-1) I(k=1) + \omega_{2i} p_{int} + u_{kint}, \quad (88)$$

$$\omega_i(h_{n1}, k) = \varpi_{0ik} + \varpi_{1k} age_n + \varpi_{2k} age_n^2 + \varpi_{3k} edu_n + \sum\nolimits_{m=1}^{9} \varpi_{ym} I(year_n = m).$$

Thus, $\omega_{2i} = (\alpha_n - \beta_n)\left[y_{f_{0t}Hk}(h_t) - y_{f_{0t}Lk}(h_t)\right] + \delta\psi_{1i}/(1-\delta)$. Simple manipulations yield that

$$\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) = y_{f_{0t}Lk}(h_t) + \beta_n \left[y_{f_{0t}Hk}(h_t) - y_{f_{0t}Lk}(h_t)\right] + \psi_{0kint} \quad (89)$$

and

$$\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) + \omega_{2i} = y_{f_{0t}Lk}(h_t) + \alpha_n \left[y_{f_{0t}Hk}(h_t) - y_{f_{0t}Lk}(h_t)\right] + \psi_{0kint} + \psi_{1i}. \quad (90)$$

From (89) and (90) it also follows that

$$\frac{\beta_n}{\alpha_n} = \frac{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) - y_{f_{0t}Lk}(h_t) - \psi_{0kint}}{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) + \omega_{2i} - y_{f_{0t}Lk}(h_t) - (\psi_{0kint} + \psi_{1i})}. \quad (91)$$

I now use (91) to derive (85). Suppose that $\psi_{1i} \leq 0$, so that the compensating wage differential is decreasing in the prior, and suppose that $y_{f_{0t}Lk}(h_t) + \psi_{0kint} \geq 0$. Thus,

$$\frac{\beta_n}{\alpha_n} \leq \frac{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) - y_{f_{0t}Lk}(h_t) - \psi_{0kint}}{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) + \omega_{2i} - y_{f_{0t}Lk}(h_t) - \psi_{0kint}} \leq \frac{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1)}{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) + \omega_{2i}}, \quad (92)$$

where the first inequality follows from $\psi_{1i} \leq 0$ and the second follows from $y_{f_{0t}Lk}(h_t) + \psi_{0kint} \geq 0$.

Next, I use (91) to derive (86). Suppose that $\psi_{1i} \geq 0$, so that the compensating wage differential is increasing in the prior, and $y_{f_{0t}Lk}(h_t) + \psi_{0kint} \leq 0$. By inverting (91), I obtain

$$\frac{\alpha_n}{\beta_n} \leq \frac{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) + \omega_{2i} - y_{f_{0t}Lk}(h_t) - \psi_{0kint}}{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) - y_{f_{0t}Lk}(h_t) - \psi_{0kint}} \leq \frac{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1) + \omega_{2i}}{\omega_i(h_{n1}, k) + \omega_{1kt} \cdot (t-1)}, \quad (93)$$

where the first inequality follows from $\psi_{1i} \geq 0$ and the second follows from $y_{f_{0t}Lk}(h_t) + \psi_{0kint} \leq 0$. This completes the proof of the claim.

5 Extended Estimation: Including Entrants at Higher Levels

Note that the probability of selection into the sample, equal to the probability of the observed assignment of a manager in the first year of employment at the firm, is determined by the model. Under the model, this probability is a function of the initial prior belief about a manager's ability, whose distribution is estimated together with the rest of the model's parameters. The assumption implicit in this formulation, and implied by the equilibrium assignment policy of the model, 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. (Specifically, these random factors reflect idiosyncratic variation in match quality and are captured by the job-specific productivity shocks.) So, with the estimation of the distribution of the initial prior, potential issues of sample selection due to the non-randomness of the data are explicitly taken into account.

However, in order to address concerns about selection possibly induced by the filtering rules I applied to the original data to obtain the estimation sample, here I report and discuss estimates of the model's parameters obtained from a sample that also contains information on managers entering the firm at levels higher than Level 1. I begin by describing this extended sample. I then turn to present the specification estimated on this sample, detailing the main differences between the specification estimated on it and that estimated on the sample of managers entering at Level 1. Finally, I discuss the estimation results based on this extended sample.

Note that in order to address selection I could have, alternatively, estimated the model on separate samples, corresponding to entrants into the firm at different levels, and compared the resulting estimates with those reported in the paper. An argument for such a choice is that all parameters governing job assignment, performance evaluations, and wages may be specific to different groups of managers as determined by their entry level. The reason I opted, instead, for one sample that includes all entrants into the firm at Levels 1–4 is to perform a clearer comparison between the parameters estimated on the sample of entrants at Level 1 and those estimated on the sample of entrants at Level 1 and higher, without relying on the added flexibility of allowing all parameters to vary across managers depending on their entry level.

5.1 Estimation Sample

Here I first describe the construction of the extended sample and then discuss the main differences between the original and extended samples in terms of job, performance, and wage patterns.

5.1.1 Sample Construction

The original BGH dataset contains 30,675 observations on entrants into one U.S. firm in a service industry between 1970 and 1979, for a total of 3,891 managers. Restricting attention to entrants at Level 1 over the period 1970–1979 leads to 21,905 observations (accounting for 71.4 percent of all observations on entrants between 1970 and 1979) and a total of 2,714 individuals.

Observe that of all individuals entering the firm at managerial levels between 1970 and 1979, 30 such entrants have missing level information, for a total of 187 employee-years. So of the 3,861 ($= 3,891 - 30$) individuals with recorded level entering into the firm between 1970 and 1979, 70.3 percent (that is, $2,714$ of $3,861$, corresponding to $(2,714/3,861) \cdot 100 = 70.3$ percent of managers) were assigned to Level 1; 29.3 percent of entrants, instead, were assigned at entry to Levels 2–4 (that is, 1,133 individuals of 3,861, corresponding to $(579+365+189)/3,861 \cdot 100 = 29.3$ percent of managers). Note that 14 managers ($= 3,861 - 2,714 - 1,133$) entered at Level 5 and higher, specifically 10 at Level 5 and 4 at Level 6. Since positions at Levels 5 and 6 correspond to top management and involve performing quite different tasks, I do not include observations on these managers in the larger sample.

Of the total 2,714 individuals entering the firm at Level 1 between 1970 and 1979, 129 managers (for a total of 283 employee-years) have missing level information at least once over their first 10 years at the firm. Deleting these individuals reduces the sample to 2,585 ($= 2,714 - 129$) managers or 20,630 employee-years. Instead, of the 1,133 individuals entering the firm at Levels 2, 3, and 4 between 1970 and 1979, overall 51 managers (for a total of 118 employee-years) have level information missing at least once over their first 10 years at the firm. Deleting these individuals reduces the sample to 1,082 ($= 1,133 - 51$) managers or 8,032 employee-years.

Of the candidate sample of 2,585 managers entering the firm at Level 1, I further restrict attention to individuals with at least 16 years of education at entry, for a total of 1,570 individuals and 10,790 employee-years. (Here 1,022 managers have between 16 and 18 years of education, and 548 managers have more than 18 years.) Of the candidate sample of 1,082 managers entering the firm at Levels 2, 3, and 4, I also restrict attention to individuals with at least 16 years of education at entry, for a total of 615 individuals and 4,236 employee-years. (Here 310 managers have between 16 and 18 years of education, and 305 managers have more than 18 years.)

Of the 1,570 managers entering the firm at Level 1 with at least 16 years of education at entry, further deleting those individuals whose recorded number of years of education changes over time reduces the sample to 1,447 individuals for a total of 9,398 employee-years. No such individual has either age or year-of-entry information missing. Of the 615 managers entering the firm at Levels 2, 3, and 4 with at least 16 years of education at entry, further deleting those individuals whose recorded number of years of education changes over time reduces the sample to 593 individuals for a total of 3,971 employee-years. Of these individuals, 319 enter at Level 2 whereas 274 enter at Levels 3 and 4. One such individual has age information missing, but none has year information missing.

Of the 1,447 entrants at Level 1, dropping the 17 individuals promoted from Level 1 to Level 3 during the first six years at the firm reduces the sample to 1,430 individuals. Of the 593 entrants at Levels 2, 3, and 4, dropping the three individuals demoted from Level 2 to Level 1 during the first six years at the firm, and one individual demoted from Level 3 to Level 2 from tenure 8 to tenure 9, reduces the sample to 589 individuals. Finally, deleting the individual with age information missing at entry reduces this latter sample to 588 individuals. Of these 588 individuals, 314 individuals entered the firm at Level 2, and 274 individuals entered at Levels 3 and 4. Finally, of the sample of 1,430 managers entering the firm at Level 1, I discard the 4 individuals with unusually high and low starting salaries whose level assignment and wage histories appear markedly different from the histories of the other managers entering at Level 1, leading to a total of 1,426 managers entering at Level 1. This is the sample I use to obtain the estimates reported in the paper. Applying a similar criterion to entrants at Levels 2, 3, and 4 leads me to discard three more individuals from the sample of managers entering the firm at Levels 2, 3, and 4, yielding a total of 585 managers entering at Levels 2, 3, and 4.

As a result, the extended estimation sample consists of 2,011 individuals corresponding to 1,426 managers entering the firm at Level 1 and 585 entering at Levels 2, 3, and 4 between 1970 and 1979 with at least 16 years of education at entry, with no level (over the first 10 years at the firm), age, education, or year-of-entry information missing, and without any change in the recorded number of years of education.

I maintain the same conventions as in the paper that observations on managers at Level 3 and higher in the data are treated as observations at job $A3$ in the model and ratings of 2, 3, 4, and 5 in the data are reclassified as ratings of zero, corresponding to low performance.

5.1.2 Differences Between Original and Extended Samples

I now discuss the salient differences between the original and extended samples. Consider the distribution of managers across levels and the associated hazard rates of separation, retention at a level, and promotion at each level in Tables A.2 and A.3. (See the corresponding Tables 1 and 2 in the paper.) Note that the proportion of managers separating from the firm at each tenure is very similar to the one in the sample of entrants into the firm at Level 1. The pattern of assignment to the other levels is also quite similar, with two main differences. First, the profile of assignment to Level 2 implies that the proportion of managers allocated to that level peaks in the second rather than the third tenure year and, past the second tenure year, the proportion of managers assigned to Level 2 is smaller than in the sample of entrants at Level 1. Second, the pattern of assignment to Level 3 in the extended sample mirrors these difference in the pattern of assignment to Level 2 across the two samples: a greater fraction of managers is assigned to Level 3 at all tenures, with, naturally, most pronounced differences at low tenures. For instance, the proportion of managers assigned to Level 3 in the original sample in the first three years of tenure is 0.0 percent in the first year, 0.0 percent in the second year, and 8.7 percent in the third year, whereas in the extended sample these proportions are 13.6, 15.6, and 23.9, respectively.

As for the implied hazard rates of job transitions, note that the hazard rates of separation, retention at a level, and promotion (to Level 2) at Level 1 in the extended sample are identical to those in the original sample of entrants at Level 1. The hazard rates of separation at Levels 2 and 3 are also very similar across the two samples. The hazard rates of retention at Level 2 and promotion from Level 2 to 3 are also strikingly similar across the two samples. The hazard rates of retention at Level 3 are quite similar too. The distributions of recorded high ratings at Levels 1 and 2 in the extended sample are also quite close to the ones in the original sample; see Table A.4.

Consider now the distribution of wages in Table A.5. By construction, the wage distribution at Level 1 is identical in the two samples. As for the distributions of wages at Levels 2 and 3, the main difference compared to the sample of entrants at Level 1 is that wages are on average higher at all tenures. This feature of the extended sample implies that individuals entering into the firm at Level

2 and higher receive on average higher wages, compared to entrants at Level 1, when assigned to the same level in the same tenure year. This evidence suggests the existence of persistent differences in productivity across managers entering into the firm at different levels. In estimation I capture these differences by allowing for differences in initial priors across managers depending on their entry levels.

5.2 Empirical Specification

Here I present the empirical specification of the model, namely, the parameterization of the processes governing initial prior beliefs, output and human capital, exogenous separations, performance ratings, and wages, respectively, for a total of 99 parameters. For each set of parameters, I discuss the differences between the specification estimated on the extended sample and that estimated on the sample of entrants at Level 1.

Initial Prior Beliefs. In specifying the distribution of initial prior beliefs, I allow for differences in this distribution across entrants into the firm at Level 1 and entrants into the firm at higher levels. I allow for this flexibility in the specification of the initial prior for two reasons. First, it provides an opportunity to validate the estimates of the parameters $\{p_{i1}\}_{i=1}^{4}$ obtained from the sample of entrants at Level 1 and discussed in the paper. Second, this formulation allows the model to better fit the larger dataset, in light of the fact that crucial parameters, like those governing the distribution of performance ratings and exogenous separations, are not allowed to vary across entrants at different levels.

Formally, I still assume that managers are one of four possible skill types, known to all model agents but unknown to the econometrician. (Here, as in the paper, I use the transformation $p_{i1} = \exp\{\phi_{i1}\}/[1+\exp\{\phi_{i1}\}]$, where ϕ_{i1} is a parameter that ranges on the real line, to avoid boundary problems in estimation.) However, I allow the value of each type's initial prior that a manager of that type is of high ability to depend on whether a manager at entry has been assigned to Level 1, resulting in the four prior parameters $\{p_{i1}\}_{i=1}^{4}$; to Level 2, resulting in the four prior parameters $\{p_{i1}\}_{i=5}^{8}$; or to Levels 3 and 4, resulting in the four prior parameters $\{p_{i1}\}_{i=9}^{12}$. Hence, the interaction between unobserved characteristics and observed level outcomes at entry leads to twelve possible 'effective' types of managers. The prior parameters for entrants at Levels 3 and 4, however, did not significantly differ from those for entrants at Level 2 across all relevant sets of parameters. For this reason, I set them equal. (That is, $p_{91} = p_{51}$, $p_{101} = p_{61}$, $p_{111} = p_{71}$, and $p_{121} = p_{81}$.) To conserve on parameters, I also maintain that $p_{71} = p_{31}$ and $p_{81} = p_{41}$, based on model diagnostics (the Akaike information criterion) and fit. Thus, estimated prior parameters are p_{11}, p_{21}, p_{31}, p_{41}, p_{51}, p_{61}, q_1, q_2, and q_3. (I also allow for interaction terms between a manager's unobserved skill type and observed entry level among the parameters of the distribution of wages; see below.)

Output and Human Capital. Omit for simplicity the subscript A for the firm in my data and the subscript n for a manager. I assume the same process for output and human capital as specified in the paper, so the human capital acquired by a manager at the firm is just a function of the manager's human capital acquired before entry into the firm, h_1, tenure at the firm, $t-1$, and previous period level assignment, k_{t-1}. As discussed above and in the paper, I normalize the parameters of expected output at the second-best firm at zero and, thus, interpret the parameters of expected output at my firm as measuring the difference between the magnitude of each such parameter at my firm and of the corresponding parameter at the second-best firm.

In light of the flexibility of the output and human capital process I specify, I conserve on parameters in several ways, following the same procedure I adopted in the paper. See the discussion in the Appendix there. The main difference between the specification estimated in the paper and the present one is as follows. Since managers entering into the firm at different levels may be differentially productive due, for instance, to their human capital acquired prior to entry into the firm, I now distinguish managers by their entry level through the variable $\ell \in \{1,2,3\}$, where $\ell = 1$ denotes entrants at Level 1 (in the data), $\ell = 2$ denotes entrants at Level 2 (in the data), and $\ell = 3$ denotes entrants at Levels 3 and 4 (in the data). Thus, an individual's state variable at the beginning of period t is now (p_{it}, ℓ, h_1, i) at $t = 1$ and $(p_{it}, \ell, h_1, t-1, k_{t-1}, i)$ at $t \geq 2$. I assume that ℓ captures the effect of h_1 on the output and human capital process at my firm. As in the paper, for the purpose of accounting for observed level assignment and separation, the only relevant dependence of expected output on i has proved to be through beliefs. Therefore, I denote the expected output at Level k at tenure t by

$$y(p_{it}, \ell, t-1, k_{t-1}, k) + \varepsilon_{kt} = b^{\ell}_{kt}(k_{t-1}) + c^{\ell}_{kt} p_{it} + \varepsilon_{kt}. \tag{94}$$

Since none of the parameters $b^{\ell}_{kt}(k_{t-1})$ significantly differed across entrants into the firm at different levels at all relevant sets of parameters, I set them equal across entry levels and simply denote them by $b_{kt}(k_{t-1})$. Adopting the same parameterizations and normalizations as in the paper, I assume that at each Level k,

$$b_{kt}(k_{t-1}) = \gamma_k(k_{t-1}) t I(t < 3) + \gamma_{kt}(k_{t-1}) I(3 \leq t \leq 7). \tag{95}$$

(In the specification in the paper, the knot in (95) is at $t = 4$.)

As a result, at Level 1 I set $b_{13}(L2) = b_{14}(L2)$ and $b_{15}(L2) = b_{16}(L2) = b_{17}(L2)$, and estimated $b_{13}(L2)$ and $b_{15}(L2)$ (in differences from the corresponding $b_{1t}(L1)$). I specified the parameters c^1_{1t} as

$$c^1_{1t} = c^1_{12} I(t = 2) + c^1_{13} I(t = 3) + c^1_{14} \sum\nolimits_{\tau=4}^{8} I(t = \tau)$$

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 at high tenures, I only estimated c^1_{12} and c^1_{13}. Since in the sample no individual is ever observed demoted, I did not estimate any slope parameter at Level 1 for entrants at higher levels ($\ell = 2, 3$). Thus, estimated parameters at Level 1 are $b_{13}(L2)$ and $b_{15}(L2)$ (in differences from the corresponding $b_{1t}(L1)$), and c^1_{12}, and c^1_{13}.

At Level 2, I assume that $c^1_{2t} = c^2_{2t}$, $2 \leq t \leq 4$ and $c^1_{26} = c^1_{27}$, $c^2_{26} = c^2_{27}$, and estimated the parameters c^1_{22}, c^1_{23}, c^1_{24}, c^1_{25}, c^2_{25}, c^1_{26}, and c^2_{26}. All other parameters are normalized to zero by the same logic as in the paper.

At Level 3, given that the hazard rates of promotion from Level 1 to 2 and from Level 2 to 3 display similar qualitative features, I allowed for common components across the parameters c_{2t}'s and c_{3t}'s at Levels 2 and 3 in order to reduce the number of parameters to estimate. In particular, proceeding as in the paper, I allow for $c^1_{3t} = \underline{c}_{3t} + c^1_{24}$ from the fourth year of tenure on, which has lead to $c^1_{34} = c^1_{35} = c^1_{24}$. I also set $c^1_{35} = c^1_{36}$, $c^3_{3t} = c^2_{3t}$ for $4 \leq t \leq 7$, and $c^1_{38} = c^2_{38} = c^3_{38}$. Differently from the specification estimated in the paper, c^ℓ_{3t}, $1 \leq t \leq 3$, did not prove significantly different from zero at all trial parameter values, so I normalized them at their values at the second-best firm to conserve on parameters. (Recall that for the sample of entrants at Level 1 analyzed in the paper, at Level 3 I just estimated the slope parameters c_{31}, c_{34}, c_{37}, and c_{38} where $c_{31} = c_{32} = c_{33}$.) Hence, estimated parameters at Level 3 are $b_{33}(L3)$, $b_{35}(L3)$, $b_{36}(L3)$, and $b_{37}(L3)$ (estimated in differences from the corresponding $b_{3t}(L2)$), and c^1_{37}, c^1_{38}, c^2_{34}, c^2_{35}, c^2_{36}, and c^2_{37}. (In the first of the two estimated specifications, I restricted $c^2_{37} = c^1_{37}$ since their difference proved insignificant.)

Note that expected output at each level in $t = 8$ is specified in the same way as in the paper.[9]

Exogenous Separations. To conserve on parameters, I assume that at Level 1 the parameters of the probabilities of exogenous separation satisfy $\eta_{11} = \eta_{12}$ and $\eta_{14} = \eta_{15} = \eta_{16} = \eta_{17} = \eta_{18}$. So estimated separation rate parameters at Level 1 are η_{11}, η_{13}, and η_{14} just as for the sample of entrants at Level 1. (More precisely, for the specification in the paper, $\eta_{11} = \eta_{12}$, $\eta_{13} = \eta_{14} + \xi_3$, and $\eta_{14} = \eta_{15} = \eta_{16} = \eta_{17} = \eta_{18}$, and estimated parameters are η_{11}, ξ_3, and η_{14}.) At Level 2, here I assume that $\eta_{21} = \eta_{22}$, $\eta_{23} = \eta_{24}$, $\eta_{25} = \eta_{26}$, and $\eta_{27} = \eta_{28}$. Then, estimated separation rate parameters at Level 2 are η_{21}, η_{23}, η_{25}, and η_{27}, whereas for the sample of entrants at Level 1, I estimate η_{21}, η_{24}, η_{25}, η_{26}, and η_{27}. (For the specification in the paper, $\eta_{21} = \eta_{22}$, $\eta_{23} = \eta_{22} + \xi_3$, and $\eta_{27} = \eta_{28}$.) At Level 3, I assume that $\eta_{33} = \eta_{34}$ and they both equal η_{31}, and $\eta_{35} = \eta_{36} = \eta_{37} = \eta_{38}$. Thus, estimated separation rate parameters at Level 3 are η_{31}, η_{32}, and η_{35}, whereas for the sample of entrants at Level 1, I estimate only η_{31}. (For the specification in the paper, $\eta_{31} = \eta_{32} = \eta_{33}$, $\eta_{3t} = \eta_{2t}$, $4 \leq t \leq 8$.)

Performance Ratings. I model the process for performance ratings here in the same way as I do in the paper. Thus, as before, estimated parameters for the true and recorded distribution of performance ratings are $\{\alpha_k, \beta_k\}^3_{k=1}$ and $(d_0, d_2(L1), d_2(L2))$.

Wages. In analogy to the specification in the paper, I assume here that at tenure t the (log) wage of manager n of skill type i, $1 \leq i \leq 12$ (as mentioned, i denotes here the 'effective type' resulting from the interaction between the unobserved skill type and the entry level of a manager), is given by

$$\ln(w^o_{Aint}) = \omega_i(h_{n1}, k) + \omega_{1t} \cdot (t-1)I(k=1) + \omega_{2i}p_{int} + u_{kint},$$

where the intercept term $\omega_i(h_{n1}, k)$ is given by

$$\omega_i(h_{n1}, k) = \varpi_{0ik} + \varpi_{1k} age_n + \varpi_{2k} age^2_n + \varpi_{3k} edu_n + \sum\nolimits_{m=1}^{9} \varpi_{ym} I(year_n = m).$$

I allow the intercept term ϖ_{0ik} to vary across managers' entry levels only when Level 3 is assigned. (This restriction amounts to $\varpi_{09k} = \varpi_{05k} = \varpi_{01k}$, $\varpi_{010k} = \varpi_{06k} = \varpi_{02k}$, $\varpi_{011k} = \varpi_{07k} = \varpi_{03k}$, and $\varpi_{012k} = \varpi_{08k} = \varpi_{04k}$ if $k = 1, 2$.) I also maintained that $\varpi_{053} = \varpi_{013}$ and $\varpi_{093} = \varpi_{013}$ since their differences proved insignificant. To avoid parameter proliferation, based on model diagnostics

[9] As in the paper, here too terms of degree higher than one in the polynomial for the match surplus value from separation, $v(p_{it}, \ell, t-1, k_{t-1}, 0)$, proved negligible.

(the Akaike information criterion) and fit, I assume that the coefficient on the prior term differs only across entrants at Level 1 and entrants at Levels 2, 3, and 4. (Equivalently, $\omega_{29} = \omega_{25}$, $\omega_{210} = \omega_{26}$, $\omega_{211} = \omega_{27}$, and $\omega_{212} = \omega_{28}$.)

I now discuss the two main differences between the specification of the process for wages estimated here and that estimated in the paper. The first difference, as just mentioned, is that here I allow the intercept term $\omega_i(h_{n1}, k)$ and the slope term ω_{2i} to vary across managers entering into the firm at different levels. The second difference is that, in light of the additional observations on wages at Level 3 in the extended sample, here I let the variance of the shock at Level 3 vary with a manager's skill type. To conserve on parameters, I assume that σ_{ik} is identical at Level k, $1 \leq k \leq 3$, for managers of the same skill type entering the firm at different levels. (Specifically, $\sigma_{9k} = \sigma_{5k} = \sigma_{1k}$, $\sigma_{10k} = \sigma_{6k} = \sigma_{2k}$, $\sigma_{11k} = \sigma_{7k} = \sigma_{3k}$, and $\sigma_{12k} = \sigma_{8k} = \sigma_{4k}$.) I also restrict $\sigma_{23} = \sigma_{13}$.

Here, as in the specification estimated in the paper, I set ϖ_{1k}, ϖ_{2k}, and ϖ_{3k}, respectively, the coefficients on age_n, age_n^2 and edu_n, equal at Levels 1 and 2. I denote their common value by ϖ_1, ϖ_2, and ϖ_3. I set $\varpi_{ym} = 0$ for $0 \leq m \leq 3$ and $\varpi_{y4} = \varpi_{y5}$, so the estimated year parameters are $\{\varpi_{ym}\}_{m=5}^{9}$, as for the specification in the paper. Here as in the paper, 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 $\omega_{15} = -\omega_{12}$. Therefore, the estimated wage parameters are $\{\varpi_{0i1}, \varpi_{0i2}, \varpi_{0i3}\}_{i=1}^{4}$, ϖ_{063}, ϖ_{073}, ϖ_{083}, ϖ_{0103}, ϖ_{0113}, ϖ_{0123}, ϖ_1, ϖ_2, ϖ_3, ϖ_{13}, ϖ_{23}, ϖ_{33}, $\{\varpi_{ym}\}_{m=5}^{9}$, ω_{12}, $\{\omega_{2i}\}_{i=1}^{8}$, $\{\sigma_{i1}, \sigma_{i2}\}_{i=1}^{4}$, σ_{13}, σ_{33}, and σ_{43}.

5.3 Estimation Results

I now discuss the results of the estimation of my model on the extended sample. I estimate two versions of the model that differ only in the specification of the error in wages at Level 3 and in one parameter normalization. Namely, in *Specification 1*, I assume that the error in wages at Level 3 is distributed according to a standard two-parameter lognormal distribution, as I assume when estimating the model on the original sample, and I maintain that $c_{37}^1 = c_{37}^2$, since their difference has proved insignificant. (Recall that c_{kt}^{ℓ} denotes the slope of expected output at job k in tenure t, averaged over productivity shocks, for managers entering into the firm at Level ℓ.) In *Specification 2*, I assume that the error in wages at Level 3 follows a more flexible three-parameter lognormal distribution.

Overall, both specifications are successful at fitting the data. (In assessing model fit for each specification, I simulated 3,000 prior realizations per manager, drawn from the estimated nonparametric distribution of initial priors.) One difference is that, being more flexible, Specification 2 fits the distribution of wages at Levels 2 and 3 better than Specification 1 and the specification in the paper.

The estimates of the main parameters of interest, namely, those governing initial uncertainty about ability, learning, and performance ratings, are remarkably similar to those reported in the paper, as discussed below. This finding, then, provides evidence that the filtering rules applied to the original data to obtain the estimation sample used in the paper have not induced appreciable selection.

5.3.1 Specification 1

I start with the fit of Specification 1 to the data. I will then discuss the main parameter estimates. Here, as in the paper, I evaluate the fit of the model by comparing observed and predicted outcomes along three dimensions: (1) the distribution of managers across levels by tenure and the hazard rates of separation, retention at a level, and promotion to the next level at each level by tenure, (2) the distribution of performance ratings at Levels 1 and 2 by tenure, and (3) the distribution of wages at each level by tenure.

Model Fit. Overall, as Tables A.2–A.5 make clear, the model estimated on the extended sample successfully captures the tenure profile of separation and assignment of the managers to the levels of the firm's hierarchy, as well as the distribution of performance ratings at Levels 1 and 2 and the wage distribution at each level and tenure. Specifically, as apparent from Table A.2, the model tracks the observed distribution of managers across levels by tenure remarkably well. In terms of the hazard rates reported in Table A.3, the model also fits well overall. Some discrepancies can be detected for the hazard rate of separation at Level 1 between the fourth and sixth years of tenure and in the hazard rate of promotion to Level 2 in the third and fifth years of tenure. The largest difference between observed and predicted outcomes at Level 2 emerges for the hazard rate of promotion between the second and third years of tenure; all other differences are modest. Instead, the hazard rates of separation and retention at Level 3 are almost perfectly matched.

Table A.4 displays the distribution of performance ratings at Levels 1 and 2 by tenure for the data and the model. The distribution of high ratings predicted by the model at each tenure tracks very closely the observed one at both levels. The main discrepancy between observed and predicted outcomes concerns the fraction of high ratings at Level 2 in the first year of tenure. One reason for this discrepancy is the small number of observations at Level 2 in this tenure year: at entry only 15 percent of managers are assigned to Level 2 while more than 70 percent are assigned to Level 1.

Finally, consider the distribution of wages by level and tenure in Table A.5 for the data and the model. Clearly, the model is quite successful at fitting these distributions, apart from Level 1 in the seventh year of tenure and Level 3 at the highest tenures. These features of model fit are analogous to those discussed in the paper.

The Pearson's χ^2 test for goodness of fit provides more favorable evidence in support of the model, not surprisingly, given the larger sample size. Specifically, in terms of the distribution of level assignments, the model is never rejected at conventional significance levels. In terms of the hazard rates of separation, retention at level, and promotion, the model is only rejected between the second and fourth years of tenure at Level 1 and between the second and third years of tenure at Level 2. In terms of the distribution of performance, the model is only rejected in the first year of tenure at Level 2. In terms of the distribution of wages, the model is only rejected at Level 1 in the first year of tenure but it is still rejected at Level 3 at most tenures.

Parameter Estimates. The loglikelihood of the sample at the estimated parameter values is 900.979, and all parameters prove significant at the 1 one percent level.[10] Given the larger sample size, almost

[10] Given the two-parameter lognormal assumption for the distribution of wages at each level, the actual wage for each

all parameters are more precisely estimated than when I estimate the model by using only observations on entrants at Level 1.

From Table A.6 a few patterns emerge. I focus here on the parameters governing initial uncertainty about ability, learning, and error in recorded performance ratings. Note that the estimated initial priors that a manager is of high ability are, respectively, $p_{11} = 0.382$, $p_{21} = 0.372$, $p_{31} = 0.466$, and $p_{41} = 0.610$ for entrants at Level 1, and by $p_{51} = 0.360$, $p_{61} = 0.400$, $p_{71} = p_{31}$, and $p_{81} = p_{41}$ for entrants at Levels 2, 3, and 4. (Recall that $p_{91} = p_{51}$, $p_{101} = p_{61}$, $p_{111} = p_{71}$, and $p_{121} = p_{81}$.) The proportions of managers of the first, second, third, and fourth skill types are given, respectively, by $q_1 = 0.102$, $q_2 = 0.290$, $q_3 = 0.360$, and $q_4 = 0.248$. According to the estimates in the paper, instead, the initial priors that a manager is of high ability for the first, second, third, and fourth skill types are given, respectively, by $p_{11} = 0.338$, $p_{21} = 0.381$, $p_{31} = 0.465$, and $p_{41} = 0.607$. There, the proportions of each such type are, respectively, $q_1 = 0.155$, $q_2 = 0.211$, $q_3 = 0.313$, and $q_4 = 0.321$.

Observe that the proportion of each type is roughly comparable across the two samples. In terms of the support of the initial priors, the main differences between the estimates obtained from the extended sample and those from the original sample concern the initial prior for managers of the first and second skill types. For the last two types, the estimates of the initial prior are almost identical across samples. Specifically, entrants at Level 1 of the first skill type are now estimated to have a higher initial prior than the one estimated on the sample of entrants at Level 1 (0.382 compared to 0.338), whereas entrants at Level 1 of the second skill type are now estimated to have a slightly lower prior (0.372 compared to 0.381). Entrants into the firm at Levels 2, 3, and 4 of both the first and second skill types are estimated to have higher initial priors than the priors for the first and second skill types estimated on the sample of entrants at Level 1 (respectively, 0.360 compared to 0.338 for the first skill type and 0.400 compared to 0.381 for the second). This finding accords with intuition: if ability is more valuable at higher levels, then managers entering into the firm at Levels 2, 3, and 4 are perceived to be more likely to be of high ability than managers entering at Level 1. Indeed, the average initial prior for entrants at Level 1 is 0.466 (from $\sum_i^I q_i p_{i1}$, compared to 0.473 estimated in the paper) with a standard deviation of 0.092 (from $[\sum_i^I q_i (p_{i1} - \sum_i^I q_i p_{i1})^2]^{1/2}$, compared to 0.102 estimated in the paper), whereas the average initial prior for entrants at Levels 2, 3, and 4 is 0.472 with a standard deviation of 0.087.

Note also that the estimates of the learning parameters are remarkably similar in magnitude and patterns to those in the paper: the estimates from the extended sample are $(\alpha_1, \beta_1) = (0.514, 0.456)$ at Level 1, $(\alpha_2, \beta_2) = (0.5432, 0.491)$ at Level 2, and $(\alpha_3, \beta_3) = (0.5429, 0.490)$ at Level 3, whereas the estimates from the sample of entrants at Level 1 are $(\alpha_1, \beta_1) = (0.514, 0.456)$ at Level 1, $(\alpha_2, \beta_2) = (0.5437, 0.491)$ at Level 2, and $(\alpha_3, \beta_3) = (0.5435, 0.490)$ at Level 3. Analogously to the estimates reported in the paper, these estimates imply that a manager of either high or low ability has the highest success rate at Level 2, the second-highest at Level 3, and the lowest at Level 1. Also, as for the estimates in the paper, here Level 1 is more informative than Level 3, which, in turn, is more informative than Level 2, since $\alpha_1\beta_3 = 0.252 > \alpha_3\beta_1 = 0.248$ and $(1-\alpha_1)(1-\beta_3) = 0.248 < (1-\alpha_3)(1-$

manager in each year is a constant that can be factored out in computing the likelihood. As in the paper, I follow this convention in reporting the likelihood value at the estimated parameter vector.

$\beta_1) = 0.249$, $\alpha_3\beta_2 = 0.267 > \alpha_2\beta_3 = 0.266$, and $(1-\alpha_3)(1-\beta_2) = 0.2327 < (1-\alpha_2)(1-\beta_3) = 0.2330$.

Finally, the fact that $d_0 = 0.487$, $d_2(L1) = -0.668$, and $d_2(L2) = -0.525$, compared to the estimates of $d_0 = 0.521$, $d_2(L1) = -0.703$, and $d_2(L2) = -0.544$ for the sample of entrants at Level 1, implies comparable estimates for the recording error in performance ratings across the two samples.

5.3.2 Specification 2

In the specification estimated in the paper and in Specification 1, wages at each level are assumed to be distributed according to a standard two-parameter lognormal distribution. In both specifications, the model does not fully capture the distribution of wages at Level 3 at high tenures. For this reason, I estimate a second, more flexible specification that allows wages at Level 3 to be distributed according to a three-parameter lognormal distribution, which, compared to the two-parameter version, features an additional location parameter.

In this new specification, Specification 2, I set the location parameter of the distribution of wages at Level 3 equal to a lower bound on managers' wages over the first eight years of tenure in the original and extended samples. More precisely, I set this lower bound at $\$20,000$ (1988 constant U.S. dollars) since the lowest observed wage is $\$20,847$. The reason for this normalization is the known computational difficulty in estimating the location parameter of a three-parameter lognormal distribution by maximum likelihood. I now turn to discuss model fit and the estimates of some of the parameters of interest.

Model Fit. I compare observed and predicted outcomes in Tables A.7–A.10. Not surprisingly, overall this specification of the model fits the data better than did Specification 1 and the one in the paper. In terms of level assignments, predicted level assignments are almost indistinguishable from the observed ones. Consider now the hazard rates of separation, retention at level, and promotion. Not surprisingly, given the small fraction of managers retained at Level 1 over time, the largest discrepancy between predicted and observed hazard rates emerges at Level 1 between the third and fourth years of tenure. At Level 2 the largest difference between observed and predicted outcomes is for the hazard rate of promotion between the second and third years of tenure. At Level 3, the predicted hazard rates are very close to the observed ones. In terms of the distribution of observed ratings here, as in the previous specification, the largest discrepancy is between the observed and predicted fraction of high ratings at Level 2 in the first year of tenure. Yet, overall, Specification 2 seems to fit the observed distribution of ratings better than Specification 1. Lastly, the model fits the distribution of wages at each level and tenure remarkably well. In particular, the fit of the distribution of wages at Levels 2 and 3 is not only substantially better than for Specification 1 but overall quite successful.

Correspondingly, the Pearson's χ^2 test provides more favorable evidence in support of the model. Specifically, in terms of the distribution of level assignments, the model is never rejected at conventional significance levels. In terms of the hazard rates of separation, retention at level, and promotion, the model is only rejected between the second and fourth years of tenure at Level 1 and between the second and third years of tenure at Level 2. In terms of the distribution of performance, the model is only rejected in the first year of tenure at Level 2. In terms of the distribution of wages, the model is only rejected at Level 1 in the first year of tenure and at Level 3 in low tenures.

Parameter Estimates. For Specification 2, the loglikelihood at the estimated parameter values is $27,904.358$, and all parameters prove significant at the 1 percent level. Consider the estimation results reported in Table A.11. For this specification, too, I confine attention to the discussion of the parameters governing initial uncertainty about ability, learning, and error in recorded performance ratings. Note that the estimated initial priors that a manager is of high ability are, respectively, $p_{11} = 0.440$, $p_{21} = 0.381$, $p_{31} = 0.465$, and $p_{41} = 0.607$ for entrants at Level 1, and by $p_{51} = 0.350$, $p_{61} = 0.400$, $p_{71} = p_{31}$, and $p_{81} = p_{41}$ for entrants at Levels 2, 3, and 4. (Recall also that $p_{91} = p_{51}$, $p_{101} = p_{61}$, $p_{111} = p_{71}$, and $p_{121} = p_{81}$.) The proportion of each type, from the first to the fourth, is given, respectively, by $q_1 = 0.123$, $q_2 = 0.284$, $q_3 = 0.317$, and $q_4 = 0.276$. Recall that the corresponding estimates in the paper are $p_{11} = 0.338$, $p_{21} = 0.381$, $p_{31} = 0.465$, and $p_{41} = 0.607$ with proportions $q_1 = 0.155$, $q_2 = 0.211$, $q_3 = 0.313$, and $q_4 = 0.321$.

In terms of the support of the initial priors, the only difference between the estimates obtained from the extended sample and those obtained from the original sample concerns the initial prior for entrants at Level 1 of the first skill type and the initial prior for entrants at higher levels of the first and second skill types. Indeed, for the last two types of managers entering at any level, the estimates of the initial prior are identical across samples. Specifically, as with Specification 1, entrants at Level 1 of the first type are now estimated to have a higher prior than that estimated on the sample of entrants at Level 1 (0.440 compared to 0.338). Here, as with Specification 1, entrants at Levels 2, 3, and 4 of both the first and second skill types are estimated to have higher priors than the priors estimated in the paper (respectively, 0.350 compared to 0.338 for the first skill type and 0.400 compared to 0.381 for the second). The proportion of each such type is quite similar across the extended and the original samples.

BGH suggest that one way to explain the difference in career paths between entrants at Level 1 and those at higher levels, typically more varied, is that unobserved abilities of new hires at higher levels vary more than those of entrants at Level 1. My estimates for Specification 2 confirm their intuition: the average initial prior for entrants at Level 1 is 0.477 (from $\sum_i^I q_i p_{i1}$, compared to 0.473 estimated in the paper) with a standard deviation of 0.087 (from $[\sum_i^I q_i(p_{i1} - \sum_i^I q_i p_{i1})^2]^{1/2}$, compared to 0.102 estimated in the paper), whereas the average initial prior for entrants at Levels 2, 3, and 4 is slightly lower, 0.472, but with a larger standard deviation of 0.091.

Finally, the estimates of the learning parameters are very similar in magnitude and patterns to those in the paper: the estimates on the extended sample are $(\alpha_1, \beta_1) = (0.514, 0.457)$ at Level 1, $(\alpha_2, \beta_2) = (0.543, 0.49059)$ at Level 2, and $(\alpha_3, \beta_3) = (0.544, 0.49058)$ at Level 3, whereas the estimates on the original sample are $(\alpha_1, \beta_1) = (0.514, 0.456)$ at Level 1, $(\alpha_2, \beta_2) = (0.5437, 0.491)$ at Level 2, and $(\alpha_3, \beta_3) = (0.5435, 0.490)$ at Level 3. Analogously to the results reported in the paper and those for Specification 1, these estimates imply that a manager of either high or low ability has the highest success rate at Level 2, the second-highest at Level 3, and the lowest at Level 1. Also, as for the estimates in the paper and those for Specification 1, here Level 1 is more informative than Level 3, which, in turn, is more informative than Level 2, since $\alpha_1\beta_3 = 0.252 > \alpha_3\beta_1 = 0.249$ and $(1 - \alpha_1)(1 - \beta_3) = 0.24758 < (1 - \alpha_3)(1 - \beta_1) = 0.24761$, $\alpha_3\beta_2 = 0.267 > \alpha_2\beta_3 = 0.266$, and

$(1-\alpha_3)(1-\beta_2) = 0.232 < (1-\alpha_2)(1-\beta_3) = 0.233$.

Finally, the fact that $d_0 = 0.507$, $d_2(L1) = -0.693$, and $d_2(L2) = -0.539$, whereas the estimates based on the original sample are $d_0 = 0.521$, $d_2(L1) = -0.703$, and $d_2(L2) = -0.544$, implies similar estimates for the classification error rates in performance ratings across the two samples.

References

Abbring, J.H. (2010): "Identification of Dynamic Discrete Choice Models," *Annual Review of Economics* 2(1), 367–394.

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.

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

Buchinsky, M., J. Hahn, and K.-il Kim (2010): "Semiparametric Information Bound of Dynamic Discrete Choice Models," *Economic Letters* 108(2), 109–112.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of the Econometrics* , Volume VIB, J.J. Heckman and E.E. Leamer Eds., Amsterdam: North-Holland, 5549–5632.

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

Fu, C. (2011): "Equilibrium Tuition, Applications, Admissions and Enrollment in the College Market," mimeo.

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

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.

Hoderlein, S., J. Klemelä, and E. Mammen (2010): "Analyzing the Random Coefficient Model Nonparametrically," *Econometric Theory* 26(3), 804–837.

Hu, Y., and M. Shum (2012): "Nonparametric Identification of Dynamic Models with Unobserved State Variables," mimeo, Johns Hopkins University, forthcoming, *Journal of Econometrics.*

Kasahara, H., and K. Shimotsu (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica* 77(1), 135–175.

Lewbel, A. (2000): "Identification of the Binary Choice Model with Misclassification," *Econometric Theory* 16(4), 603–609.

Magnac, T., and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica* 70(2), 801–816.

Meijer, E., and J.Y. Ypma (2008): "A Simple Identification Proof for a Mixture of Two Univariate Normal Distributions," *Journal of Classification* 25(1), 113–123.

Pastorino, E., and P.J. Kehoe (2012): "Efficiency and Pricing in Markets with Experience Goods", mimeo.

Pakes, A., and R. Ericson (1998): "Empirical Implications of Alternative Models of Firm Dynamics," *Journal of Economic Theory* 79(1), 1–45.

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

— (1988): " Maximum Likelihood Estimation of Discrete Control Processes," *SIAM Journal on Control and Optimization* 26(5), 1006–1023.

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

Figure 1. Static Expected Output and Statically Optimal Policies

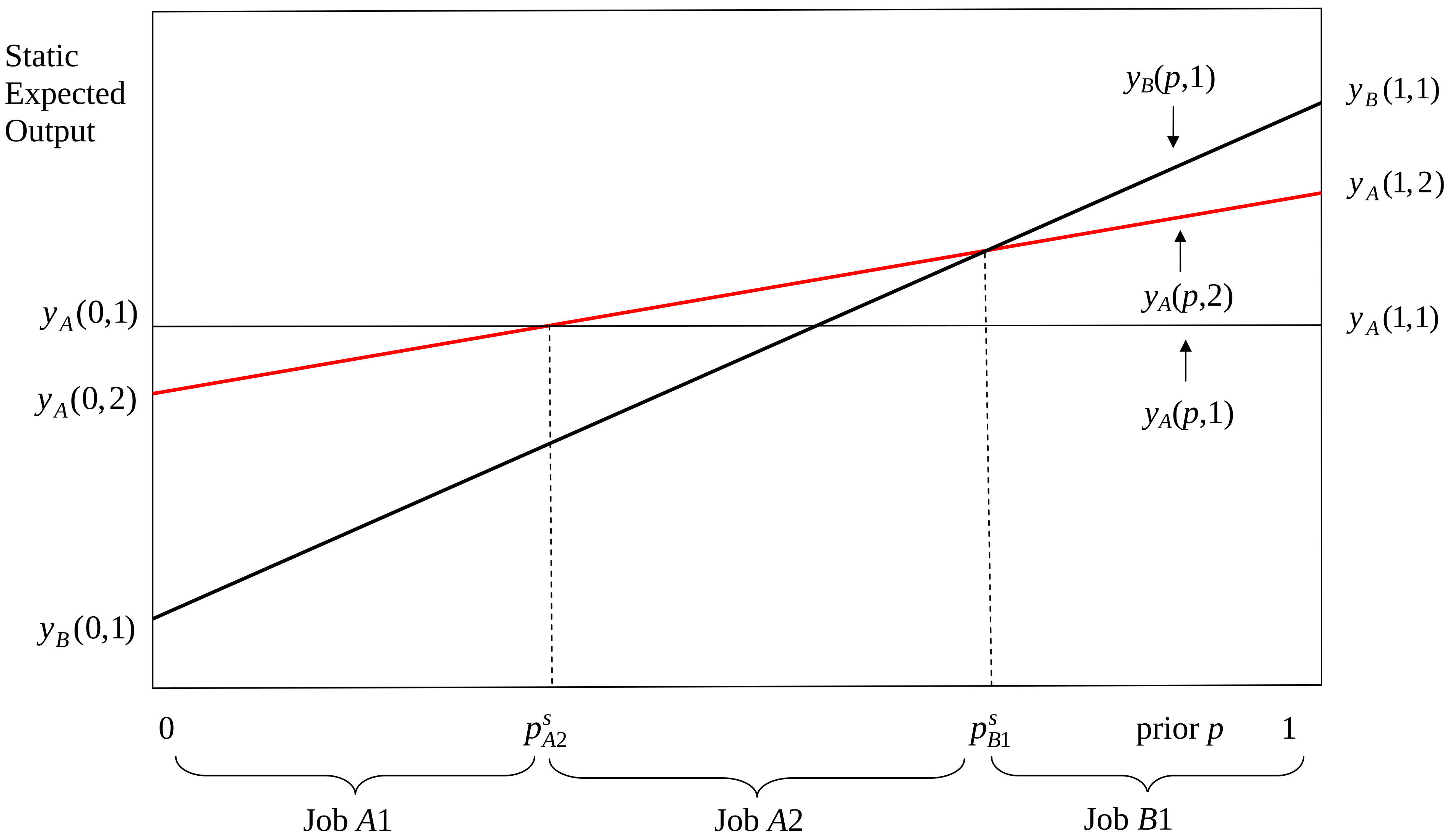
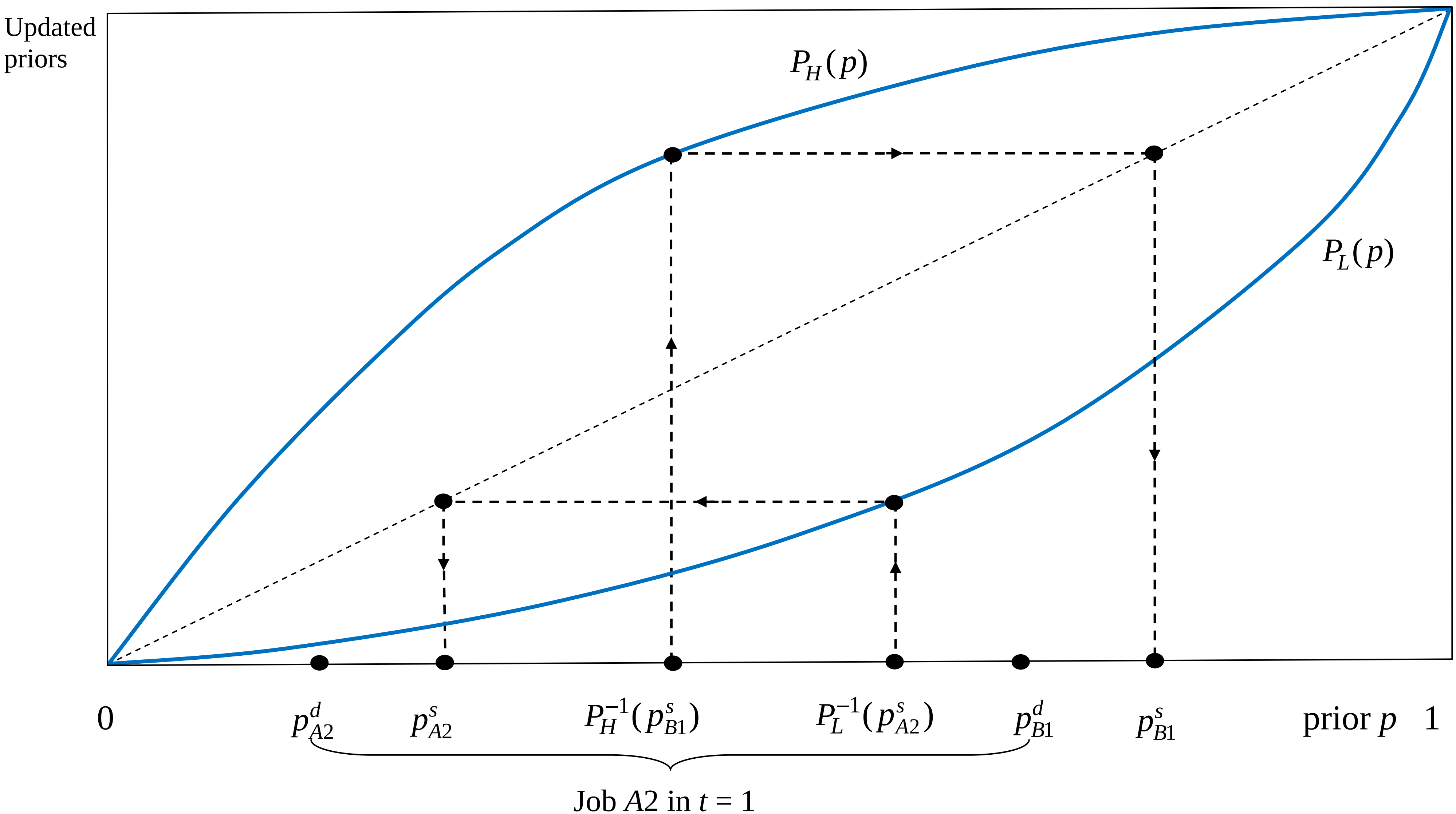


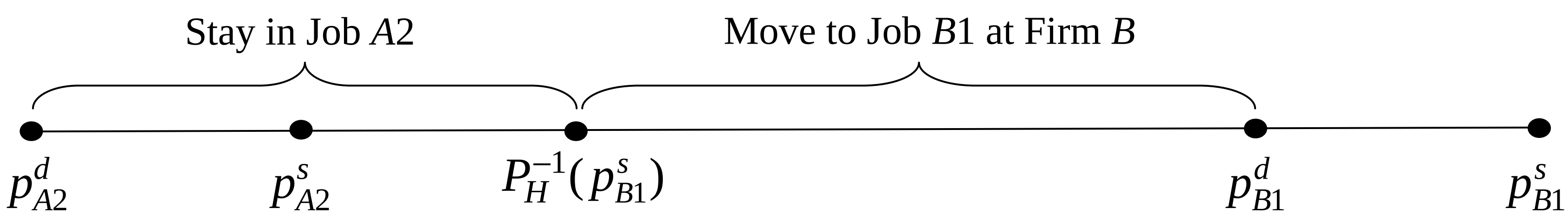
Figure 2. Bayesian Updating in Job $A2$



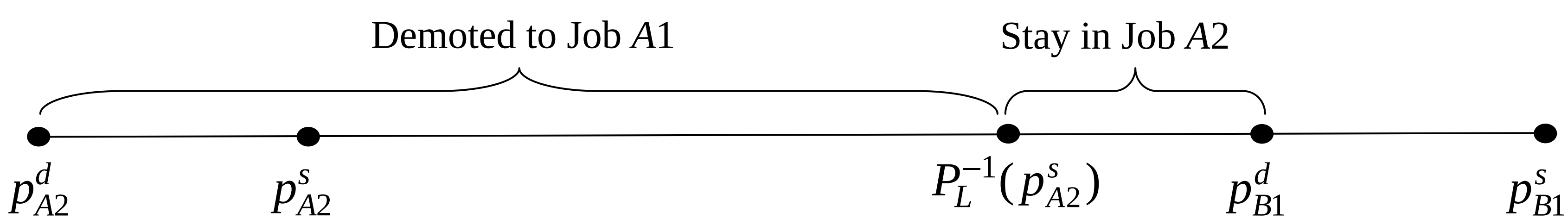
*Note three conditions hold: $P_H(p^d_{A2}) < p^s_{B1}$, $P_L(p^d_{B1}) \geq p^s_{A2}$, and $P_H(p^d_{B1}) \geq p^s_{B1}$

Figure 3. Jobs Assigned in Period 2 After Job *A*2 in Period 1*

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



B. Period 2 Assignment After Failure in Job *A*2 in Period 1



*All workers with priors between p_{A2}^{d} and p_{B1}^{d} work in job *A*2 at firm *A* in period 1

TABLE A.1
Results from Monte Carlo Simulations

Parameters	Baseline Estimates	Bias	t-Statistic of Bias	St. Dev. of Estimated Parameter	Mean of Estimated St. Error
Prior Distribution					
ϕ_{11} (p_{11} = 0.338)	–0.672	–0.017	–3.428	0.035	0.069
ϕ_{21} (p_{21} = 0.381)	–0.484	0.002	0.416	0.042	0.003
ϕ_{31} (p_{31} = 0.465)	–0.141	0.005	1.392	0.027	0.001
ϕ_{41} (p_{41} = 0.607)	0.435	0.032	5.117	0.045	0.003
q_1	0.155	–0.021	–9.613	0.016	0.019
q_2	0.211	–0.015	–7.200	0.015	0.001
q_3	0.313	0.010	2.755	0.026	0.002
Probability of High Output					
α_1	0.514	0.0001	0.443	0.002	0.026
β_1	0.456	–0.0003	–1.249	0.002	0.008
α_2	0.5437	–0.0002	–1.756	0.001	0.004
β_2	0.491	0.00001	0.103	0.001	0.013
α_3	0.5435	–0.0003	–2.205	0.001	0.001
β_3	0.490	0.0001	0.720	0.001	0.019
Ratings Error					
d_0	0.521	–0.007	–0.645	0.079	0.007
$d_2(L1)$	–0.703	0.003	0.395	0.048	0.001
$d_2(L2)$	–0.544	–0.001	–0.122	0.033	0.033
Output and Human Capital					
$b_{14}(L2)$ - $b_{14}(L1)$	–704.735	–0.837	–0.515	11.493	3.166
$b_{15}(L2)$ - $b_{15}(L1)$	–479.607	–0.304	–0.609	3.535	2.823
c_{12}	2,960.515	–3.752	–1.078	24.612	46.467
c_{22}	1,858.714	0.381	0.221	12.185	1.930
c_{23}	1,505.367	4.036	1.961	14.554	33.448
c_{25}	1,629.309	3.025	1.375	15.555	8.282
c_{26}	1,745.184	2.476	1.623	10.786	1.959
$b_{34}(L3)$ - $b_{34}(L2)$	853.477	1.874	0.588	22.518	16.670
$b_{35}(L3)$ - $b_{35}(L2)$	202.791	1.115	1.200	6.570	0.693
$b_{37}(L3)$ - $b_{37}(L2)$	228.069	2.124	1.325	11.334	2.620
c_{31}	–399.955	–1.758	–0.885	14.047	10.633
c_{34}	2,963.404	1.87	0.527	16.739	13.532
c_{37}	2,190.704	0.541	0.326	11.724	1.112
c_{38}	2,003.340	4.183	2.769	10.681	232.904
Exogenous Separation					
η_{11}	0.145	–0.0001	–0.721	0.001	0.001
ξ_3	0.033	–0.0004	–2.597	0.001	0.0002
η_{14}	0.050	0.0001	1.830	0.0002	0.0001
η_{21}	0.136	0.0001	0.592	0.001	0.004
η_{24}	0.142	–0.0002	–2.796	0.001	0.0001
η_{25}	0.121	0.0001	1.253	0.001	0.027
η_{26}	0.115	–0.0002	–1.596	0.001	0.007
η_{27}	0.111	0.0002	1.808	0.001	0.0003
η_{31}	0.122	–0.0002	–1.498	0.001	0.001

TABLE A.1 (Continued)
Results from Monte Carlo Simulations

Parameters	Baseline Estimates	Bias	t-Statistic of Bias	St. Dev. of Estimated Parameter	Mean of Estimated St. Error
Parameters of $\omega_{ik}(age,edu,year)$					
ϖ_{011}	8.805	–0.025	–7.553	0.024	0.145
ϖ_{021}	9.288	0.042	13.374	0.022	0.003
ϖ_{031}	9.213	0.019	8.559	0.015	0.005
ϖ_{041}	8.865	0.007	3.328	0.015	0.009
ϖ_{012}	8.969	–0.026	–7.148	0.026	0.061
ϖ_{022}	9.359	0.044	14.110	0.022	0.002
ϖ_{032}	9.281	0.022	11.035	0.014	0.004
ϖ_{042}	8.945	0.010	4.855	0.015	0.006
ϖ_{013}	9.534	–0.019	–3.785	0.035	0.002
ϖ_{023}	9.813	0.050	11.745	0.030	0.002
ϖ_{033}	9.738	0.030	7.684	0.028	0.004
ϖ_{043}	9.418	0.003	0.711	0.026	0.006
ϖ_{1}	0.028	0.0003	4.928	0.0004	0.001
ϖ_{2}	–0.0003	–0.000004	–3.800	0.00001	0.000001
ϖ_{3}	0.022	0.0003	2.533	0.001	0.00001
ϖ_{13}	0.010	0.0003	3.855	0.001	0.0001
ϖ_{23}	–0.0001	–0.000004	–2.630	0.00001	0.000002
ϖ_{33}	0.021	0.0003	1.600	0.001	0.0002
ϖ_{y5}	–0.063	0.009	5.509	0.011	0.0003
ϖ_{y6}	–0.107	0.002	1.167	0.012	0.002
ϖ_{y7}	–0.140	–0.003	–1.303	0.014	0.002
ϖ_{y8}	–0.208	–0.005	–3.244	0.011	0.002
ϖ_{y9}	–0.169	0.002	1.128	0.012	0.001
Coefficient on Tenure					
ω_{12}	0.007	–0.0004	–2.942	0.001	0.0001
Coefficients on Prior by Type					
ω_{21}	2.371	0.006	0.644	0.068	0.010
ω_{22}	1.833	–0.098	–13.330	0.052	0.007
ω_{23}	1.316	–0.054	–11.105	0.034	0.005
ω_{24}	1.364	–0.042	–9.473	0.031	0.003
Wage Standard Deviations by Type and Level					
σ_{11}	0.076	–0.017	–29.924	0.004	0.00002
σ_{21}	0.070	–0.009	–17.459	0.004	0.001
σ_{31}	0.057	–0.008	–17.998	0.003	0.001
σ_{41}	0.044	–0.006	–23.703	0.002	0.0004
σ_{12}	0.063	–0.021	–41.064	0.004	0.00001
σ_{22}	0.047	–0.014	–41.553	0.002	0.0004
σ_{32}	0.0302	–0.008	–38.472	0.002	0.0002
σ_{42}	0.0303	–0.008	–34.107	0.002	0.0002
σ_{3}	0.047	–0.021	–146.134	0.001	0.00004

TABLE A.2
Percentage Distribution of Managers Across Levels by Tenure
(Extended Sample Specification 1)

	Separation		Level 1		Level 2		Level 3	
Tenure	Data	Model	Data	Model	Data	Model	Data	Model
1	0.0	0.0	70.9	70.8	15.5	15.6	13.6	13.6
2	14.4	14.9	32.4	32.1	37.6	37.7	15.6	15.3
3	27.3	28.8	11.9	11.7	36.9	36.4	23.9	23.2
4	37.5	38.5	5.5	5.2	22.5	22.4	34.6	33.8
5	46.1	46.4	3.3	3.3	13.9	14.2	36.7	36.2
6	52.1	52.8	2.0	2.2	9.4	9.5	36.4	35.5
7	57.6	58.5	1.5	1.6	6.0	6.3	34.9	33.6

TABLE A.3
Hazard Rates of Separation, Retention at Level, and Promotion (Percentages)
(Extended Sample Specification 1)

		Separated		Retained		Promoted	
Level	Tenure	Data	Model	Data	Model	Data	Model
Level 1	1 to 2	14.4	14.4	45.7	45.3	39.9	40.3
	2 to 3	14.6	14.4	36.9	36.4	48.5	41.8
	3 to 4	12.1	13.6	45.8	44.7	42.1	26.2
	4 to 5	11.8	5.4	60.0	62.2	28.2	21.7
	5 to 6	9.1	5.4	62.1	66.5	28.8	20.1
	6 to 7	12.2	5.4	73.2	74.0	14.6	14.7
Level 2	1 to 2	16.3	19.0	60.3	59.2	23.4	21.8
	2 to 3	16.1	19.0	56.3	60.9	27.6	20.1
	3 to 4	15.8	14.4	47.3	53.2	36.9	32.4
	4 to 5	15.9	14.4	54.9	58.0	29.2	27.6
	5 to 6	13.3	13.0	60.9	62.4	25.8	24.6
	6 to 7	14.3	12.9	60.8	62.8	24.9	24.3
Level 3	1 to 2	12.1	12.7	87.9	87.3		
	2 to 3	13.1	13.9	86.9	86.1		
	3 to 4	12.5	12.7	87.5	87.3		
	4 to 5	12.7	12.7	87.3	87.1		
	5 to 6	10.6	12.2	89.4	87.7		
	6 to 7	10.6	12.2	89.4	87.8		

TABLE A.4
Percentage of High Ratings at Levels 1 and 2
(Extended Sample Specification 1)

	Level 1		Level 2	
Tenure	Data	Model	Data	Model
1	52.7	51.3	58.8	67.3
2	34.9	35.3	56.1	55.2
3	20.0	22.1	42.8	42.4
4	11.8	12.7	26.0	30.5
5	2.4	7.1	17.7	20.6
6	3.7	3.7	11.3	13.3
7	0.0	1.9	12.9	8.4

TABLE A.5
Percentage Wage Distributions by Level and Tenure
(Extended Sample Specification 1)

		Between $20K and $40K		Between $40K and $60K		Between $60K and $80K	
Level	Tenure	Data	Model	Data	Model	Data	Model
Level 1	1	59.1	55.6	40.5	43.7	0.4	0.7
	2	54.5	55.6	44.7	43.5	0.8	0.9
	3	55.8	56.6	44.2	42.2	0.0	1.3
	4	54.2	55.7	45.8	42.7	0.0	1.5
	5	64.1	65.9	35.9	33.1	0.0	0.9
	6	69.2	66.7	30.8	32.1	0.0	1.0
	7	75.0	68.0	25.0	30.7	0.0	1.1
Level 2	1	13.3	12.5	67.7	69.5	18.7	17.7
	2	29.0	28.5	65.6	65.5	5.4	5.9
	3	29.4	33.0	66.3	63.4	4.3	3.6
	4	34.7	35.4	60.5	61.1	4.8	3.5
	5	35.6	36.7	60.6	59.7	3.8	3.5
	6	40.7	37.6	54.8	58.6	4.5	3.7
	7	38.9	38.2	58.4	57.9	2.7	3.8
Level 3	1	6.9	3.7	36.8	31.5	48.9	35.1
	2	5.3	4.7	45.7	40.7	43.1	34.1
	3	4.1	6.0	64.2	59.7	30.2	25.4
	4	4.5	8.3	72.4	68.2	22.4	18.8
	5	4.2	9.4	74.9	69.6	20.1	17.1
	6	5.5	10.4	77.7	69.3	16.5	16.9
	7	3.7	11.3	77.5	68.7	18.8	16.7

TABLE A.6
Estimates of Model Parameters (Extended Sample Specification 1)

Parameters	Value	Asymptotic Standard Error
Prior Distribution		
ϕ_{11} (p_{11} = 0.382)	–0.480	0.033
ϕ_{21} (p_{21} = 0.372)	–0.525	0.027
ϕ_{31} (p_{31} = 0.466)	–0.138	0.018
ϕ_{41} (p_{41} = 0.610)	0.447	0.026
ϕ_{51} (p_{51} = 0.360)	–0.575	0.048
ϕ_{61} (p_{61} = 0.400)	–0.405	0.036
q_1	0.102	0.009
q_2	0.290	0.045
q_3	0.360	0.092
Probability of High Output		
α_1	0.514	0.032
β_1	0.456	0.012
α_2	0.5432	0.003
β_2	0.491	0.068
α_3	0.5429	0.007
β_3	0.490	0.010
Ratings Error		
d_0	0.487	0.037
$d_2(L1)$	–0.668	0.037
$d_2(L2)$	–0.525	0.028
Output and Human Capital		
c_{12}^1	2,476.092	16.758
$b_{13}(L2)$ - $b_{13}(L1)$	–705.921	5.452
c_{13}^1	2,419.015	8.489
$b_{15}(L2)$ - $b_{15}(L1)$	–1,092.881	1.880
c_{22}^1	2,692.672	8.209
c_{23}^1	1,800.283	2.316
c_{24}^1	1,860.990	1.612
c_{25}^1	1,275.078	2.297
c_{25}^2	8,145.406	1.493
c_{26}^1	1,546.718	1.536
c_{26}^2	2,664.069	1.015
$b_{33}(L3)$ - $b_{33}(L2)$	1,267.237	1.660
c_{34}^2	1,789.365	2.008
$b_{35}(L3)$ - $b_{35}(L2)$	153.565	3.918
c_{35}^2	7,631.714	2.325
$b_{36}(L3)$ - $b_{36}(L2)$	180.222	2.053
c_{36}^2	1,879.019	3.449
$b_{37}(L3)$ - $b_{37}(L2)$	327.530	2.195
c_{37}^1	2,615.709	36.094
c_{38}^1	2,054.190	0.292

TABLE A.6 (Continued)
Estimates of Model Parameters (Extended Sample Specification 1)

Parameters	Value	Asymptotic Standard Error
Exogenous Separation		
η_{11}	0.144	0.004
η_{13}	0.136	0.002
η_{14}	0.054	0.0001
η_{21}	0.190	0.003
η_{23}	0.144	0.001
η_{25}	0.129	0.0003
η_{27}	0.123	0.0003
η_{31}	0.127	0.001
η_{32}	0.139	0.001
η_{35}	0.122	0.0003
Parameters of $\omega_{ik}(age,edu,year)$		
ϖ_{011}	8.431	0.007
ϖ_{021}	9.217	0.004
ϖ_{031}	9.055	0.006
ϖ_{041}	8.758	0.010
ϖ_{012}	8.589	0.006
ϖ_{022}	9.283	0.004
ϖ_{032}	9.143	0.005
ϖ_{042}	8.845	0.007
ϖ_{013}	9.168	0.006
ϖ_{023}	9.773	0.006
ϖ_{033}	9.647	0.007
ϖ_{043}	9.377	0.013
ϖ_{063}	9.735	0.010
ϖ_{073}	9.627	0.009
ϖ_{083}	9.404	0.019
ϖ_{0103}	9.852	0.004
ϖ_{0113}	9.692	0.005
ϖ_{0123}	10.095	0.007
ϖ_{1}	0.037	0.0001
ϖ_{2}	–0.0004	0.000002
ϖ_{3}	0.018	0.0005
ϖ_{13}	0.017	0.0003
ϖ_{23}	–0.0002	0.000005
ϖ_{33}	0.016	0.001
ϖ_{y5}	–0.062	0.003
ϖ_{y6}	–0.147	0.004
ϖ_{y7}	–0.151	0.003
ϖ_{y8}	–0.215	0.003
ϖ_{y9}	–0.152	0.003

TABLE A.6 (Continued)
Estimates of Model Parameters (Extended Sample Specification 1)

Parameters	Value	Asymptotic Standard Error
Coefficient on Tenure		
ω_{12}	0.007	0.0003
Coefficients on Prior by Type		
ω_{21}	2.674	0.060
ω_{22}	1.758	0.038
ω_{23}	1.359	0.018
ω_{24}	1.335	0.015
ω_{25}	2.604	0.091
ω_{26}	2.001	0.049
ω_{27}	1.598	0.020
ω_{28}	1.486	0.016
Wage Standard Deviations by Type and Level		
σ_{11}	0.077	0.001
σ_{21}	0.070	0.001
σ_{31}	0.056	0.001
σ_{41}	0.041	0.001
σ_{12}	0.079	0.001
σ_{22}	0.057	0.001
σ_{32}	0.044	0.0004
σ_{42}	0.037	0.0004
σ_{13}	0.081	0.0005
σ_{33}	0.048	0.0005
σ_{43}	0.091	0.001

TABLE A.7
Percentage Distribution of Managers Across Levels by Tenure
(Extended Sample Specification 2)

	Separation		Level 1		Level 2		Level 3	
Tenure	Data	Model	Data	Model	Data	Model	Data	Model
1	0.0	0.0	70.9	70.8	15.5	15.6	13.6	13.6
2	14.4	14.3	32.4	32.4	37.6	38.1	15.6	15.2
3	27.3	27.8	11.9	12.0	36.9	36.9	23.9	23.4
4	37.5	38.0	5.5	5.4	22.5	22.4	34.6	34.2
5	46.1	46.2	3.3	3.3	13.9	14.1	36.7	36.4
6	52.1	52.5	2.0	2.2	9.4	9.5	36.4	35.7
7	57.6	58.1	1.5	1.7	6.0	6.2	34.9	33.9

TABLE A.8
Hazard Rates of Separation, Retention at Level, and Promotion (Percentages)
(Extended Sample Specification 2)

		Separated		Retained		Promoted	
Level	Tenure	Data	Model	Data	Model	Data	Model
Level 1	1 to 2	14.4	13.6	45.7	45.8	39.9	40.3
	2 to 3	14.6	13.6	36.9	36.9	48.5	41.0
	3 to 4	12.1	13.2	45.8	44.9	42.1	26.5
	4 to 5	11.8	5.3	60.0	62.3	28.2	21.8
	5 to 6	9.1	5.3	62.1	66.4	28.8	20.4
	6 to 7	12.2	5.2	73.2	77.1	14.6	12.7
Level 2	1 to 2	16.3	18.1	60.3	61.4	23.4	20.4
	2 to 3	16.1	18.1	56.3	61.8	27.6	20.0
	3 to 4	15.8	14.7	47.3	52.2	36.9	33.0
	4 to 5	15.9	14.7	54.9	57.6	29.2	27.6
	5 to 6	13.3	13.2	60.9	62.6	25.8	24.1
	6 to 7	14.3	13.1	60.8	62.5	24.9	24.3
Level 3	1 to 2	12.1	13.4	87.9	86.6		
	2 to 3	13.1	14.6	86.9	85.4		
	3 to 4	12.5	13.5	87.5	86.5		
	4 to 5	12.7	13.4	87.3	86.6		
	5 to 6	10.6	11.8	89.4	88.1		
	6 to 7	10.6	11.8	89.4	88.2		

TABLE A.9
Percentage of High Ratings at Levels 1 and 2
(Extended Sample Specification 2)

	Level 1		Level 2	
Tenure	Data	Model	Data	Model
1	52.7	51.5	58.8	67.9
2	34.9	34.9	56.1	55.6
3	20.0	21.4	42.8	42.5
4	11.8	12.0	26.0	30.3
5	2.4	6.5	17.7	20.3
6	3.7	3.4	11.3	13.0
7	0.0	1.7	12.9	8.0

TABLE A.10
Percentage Wage Distributions by Level and Tenure
(Extended Sample Specification 2)

		Between $20K and $40K		Between $40K and $60K		Between $60K and $80K	
Level	Tenure	Data	Model	Data	Model	Data	Model
Level 1	1	59.1	55.2	40.5	44.2	0.4	0.6
	2	54.5	55.2	44.7	43.9	0.8	0.9
	3	55.8	57.2	44.2	41.4	0.0	1.4
	4	54.2	56.8	45.8	41.5	0.0	1.7
	5	64.1	67.4	35.9	31.6	0.0	0.9
	6	69.2	68.7	30.8	30.1	0.0	1.0
	7	75.0	69.6	25.0	29.0	0.0	1.0
Level 2	1	13.3	13.6	67.7	68.9	18.7	17.3
	2	29.0	27.5	65.6	66.6	5.4	5.8
	3	29.4	32.4	66.3	63.8	4.3	3.8
	4	34.7	35.0	60.5	61.2	4.8	3.8
	5	35.6	35.8	60.6	60.3	3.8	3.8
	6	40.7	36.4	54.8	59.2	4.5	4.3
	7	38.9	37.0	58.4	58.5	2.7	4.4
Level 3	1	6.9	0.8	36.8	32.7	48.9	35.8
	2	5.3	1.6	45.7	43.1	43.1	34.1
	3	4.1	2.3	64.2	64.8	30.2	24.6
	4	4.5	3.7	72.4	75.1	22.4	17.0
	5	4.2	4.3	74.9	76.9	20.1	15.5
	6	5.5	4.9	77.7	77.5	16.5	14.7
	7	3.7	5.4	77.5	77.1	18.8	14.7

TABLE A.11
Estimates of Model Parameters (Extended Sample Specification 2)

Parameters	Value	Asymptotic Standard Error
Prior Distribution		
ϕ_{11} ($p_{11} = 0.440$)	–0.242	0.026
ϕ_{21} ($p_{21} = 0.381$)	–0.486	0.034
ϕ_{31} ($p_{31} = 0.465$)	–0.140	0.018
ϕ_{41} ($p_{41} = 0.607$)	0.433	0.026
ϕ_{51} ($p_{51} = 0.350$)	–0.618	0.036
ϕ_{61} ($p_{61} = 0.400$)	–0.407	0.032
q_1	0.123	0.011
q_2	0.284	0.041
q_3	0.317	0.058
Probability of High Output		
α_1	0.514	0.065
β_1	0.457	0.011
α_2	0.543	0.006
β_2	0.49059	0.013
α_3	0.544	0.007
β_3	0.49058	0.010
Ratings Error		
d_0	0.507	0.037
$d_2(L1)$	–0.693	0.037
$d_2(L2)$	–0.539	0.028
Output and Human Capital		
c_{12}^{1}	2,546.379	60.027
$b_{13}(L2)$ - $b_{13}(L1)$	–921.611	44.749
c_{13}^{1}	2,654.591	39.086
$b_{15}(L2)$ - $b_{15}(L1)$	–1,222.480	42.863
c_{22}^{1}	2,528.019	47.190
c_{23}^{1}	1,981.858	118.413
c_{24}^{1}	1,749.285	5.412
c_{25}^{1}	1,228.735	7.245
c_{25}^{2}	4,315.857	29.517
c_{26}^{1}	1,511.991	20.117
c_{26}^{2}	5,074.299	71.326
$b_{33}(L3)$ - $b_{33}(L2)$	1,586.090	22.701
c_{34}^{2}	1,713.867	14.304
$b_{35}(L3)$ - $b_{35}(L2)$	352.196	50.346
c_{35}^{2}	3,544.000	645.993
$b_{36}(L3)$ - $b_{36}(L2)$	182.064	0.047
c_{36}^{2}	4,046.057	30.651
$b_{37}(L3)$ - $b_{37}(L2)$	322.609	1.695
c_{37}^{1}	2,463.459	9.025
c_{37}^{2}	4,727.139	7.344
c_{38}^{1}	2,025.335	108.838

TABLE A.11 (Continued)
Estimates of Model Parameters (Extended Sample Specification 2)

Parameters	Value	Asymptotic Standard Error
Exogenous Separation		
η_{11}	0.136	0.003
η_{13}	0.133	0.002
η_{14}	0.053	0.0001
η_{21}	0.181	0.002
η_{23}	0.148	0.001
η_{25}	0.132	0.0003
η_{27}	0.126	0.0003
η_{31}	0.134	0.001
η_{32}	0.146	0.001
η_{35}	0.118	0.0004
Parameters of $\omega_{ik}(age,edu,year)$		
ϖ_{011}	8.398	0.006
ϖ_{021}	9.275	0.005
ϖ_{031}	9.140	0.008
ϖ_{041}	8.904	0.010
ϖ_{012}	8.559	0.007
ϖ_{022}	9.340	0.004
ϖ_{032}	9.217	0.006
ϖ_{042}	8.990	0.008
ϖ_{013}	8.511	0.008
ϖ_{023}	9.317	0.008
ϖ_{033}	9.136	0.009
ϖ_{043}	8.849	0.013
ϖ_{063}	9.333	0.013
ϖ_{073}	9.172	0.011
ϖ_{083}	9.041	0.024
ϖ_{0103}	9.508	0.008
ϖ_{0113}	9.275	0.006
ϖ_{0123}	9.904	0.009
ϖ_{1}	0.036	0.0001
ϖ_{2}	–0.0004	0.000002
ϖ_{3}	0.012	0.0005
ϖ_{13}	0.010	0.001
ϖ_{23}	–0.0001	0.00002
ϖ_{33}	0.018	0.001
ϖ_{y5}	–0.062	0.003
ϖ_{y6}	–0.114	0.004
ϖ_{y7}	–0.151	0.004
ϖ_{y8}	–0.222	0.003
ϖ_{y9}	–0.162	0.003

TABLE A.11 (Continued)
Estimates of Model Parameters (Extended Sample Specification 2)

Parameters	Value	Asymptotic Standard Error
Coefficient on Tenure		
ω_{12}	0.008	0.0003
Coefficients on Prior by Type		
ω_{21}	2.767	0.040
ω_{22}	1.917	0.044
ω_{23}	1.487	0.018
ω_{24}	1.324	0.014
ω_{25}	3.228	0.083
ω_{26}	2.196	0.046
ω_{27}	1.747	0.020
ω_{28}	1.507	0.015
Wage Standard Deviations by Type and Level		
σ_{11}	0.087	0.001
σ_{21}	0.069	0.001
σ_{31}	0.057	0.001
σ_{41}	0.043	0.001
σ_{12}	0.081	0.001
σ_{22}	0.057	0.001
σ_{32}	0.038	0.0005
σ_{42}	0.036	0.0004
σ_{13}	0.112	0.001
σ_{33}	0.098	0.001
σ_{43}	0.135	0.002

TABLE A.12
Counterfactual Experiments: Importance of Experimentation for Wages
Baseline, Equal Informativeness as Levels 1, 2, and 3*

	Wages in Each Case			
		Equal Informativeness as		
Statistic	Baseline	Level 1	Level 2	Level 3
Means by Level				
Level 1	$39,584	$39,763	$39,785	$39,791
Level 2	43,179	42,600	43,031	43,027
Level 3	48,963	48,818	48,874	48,881
Standard Deviations by Level				
Level 1	$6,936	$6,902	$6,942	$6,945
Level 2	7,077	6,831	7,094	7,106
Level 3	8,046	7,971	7,890	7,914
Cumulative Growth Rates				
Tenure 2	4.6%	0.9%	0.9%	0.9%
Tenure 3	8.9	17.6	9.3	9.3
Tenure 4	13.8	20.5	14.3	14.3
Tenure 5	15.9	21.6	16.6	16.6
Tenure 6	17.5	22.1	18.3	18.2
Tenure 7	18.5	22.2	19.2	19.2
Tenure 7 (Balanced Panel)	19.4	23.3	20.2	20.1

*Equal Info as Level: 1, $\alpha_k = \hat{\alpha}_1$, $\beta_k = \hat{\beta}_1$, $k = 2,3$; 2, $\alpha_k = \hat{\alpha}_2$, $\beta_k = \hat{\beta}_2$, $k = 1,3$; 3, $\alpha_k = \hat{\alpha}_3$, $\beta_k = \hat{\beta}_3$, $k = 1,2$

TABLE A.13
Counterfactual Experiment: Importance of Experimentation for Level Assignments
Baseline and Equal Informativeness as Level 2*

	Separation		Level 1		Level 2		Level 3	
Tenure	Base.	Equal Info. As L2	Base.	Equal Info. As L2	Base.	Equal Info. As L2	Base.	Equal Info. As L2
1	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0
2	14.5	14.5	45.7	84.8	39.8	0.6	0.0	0.0
3	26.5	26.9	17.2	14.6	47.3	49.1	8.9	9.3
4	37.1	37.6	8.1	6.0	29.2	29.7	25.6	26.7
5	45.3	45.9	5.3	3.6	18.3	18.1	31.2	32.4
6	51.5	52.2	3.4	2.2	12.6	12.2	32.5	33.4
7	56.9	57.5	2.7	1.7	8.3	7.9	32.1	32.9

*Equal Informativeness as Level 2: $\alpha_k = \hat{\alpha}_2,\ \beta_k = \hat{\beta}_2,\ k = 1, 3$

TABLE A.14
Counterfactual Experiment: Importance of Experimentation for Level Assignments
Baseline and Equal Informativeness as Level 3*

	Separation		Level 1		Level 2		Level 3	
Tenure	Base.	Equal Info. As L3	Base.	Equal Info. As L3	Base.	Equal Info. As L3	Base.	Equal Info. As L3
1	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0
2	14.5	14.5	45.7	84.8	39.8	0.6	0.0	0.0
3	26.5	26.9	17.2	15.3	47.3	48.1	8.9	9.7
4	37.1	37.5	8.1	6.5	29.2	29.1	25.6	26.9
5	45.3	45.8	5.3	4.0	18.3	17.8	31.2	32.4
6	51.5	52.1	3.4	2.5	12.6	12.0	32.5	33.4
7	56.9	57.4	2.7	1.9	8.3	7.8	32.1	32.8

*Equal Informativeness as Level 3: $\alpha_k = \hat{\alpha}_3,\ \beta_k = \hat{\beta}_3,\ k = 1, 2$

TABLE A.15
Percentage of High Ratings at Level and Conditional on Retention at Level or Promotion

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