Federal Reserve Bank of Minneapolis
Research Department Staff Report 470

July 2012

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

Elena Pastorino

Federal Reserve Bank of Minneapolis
and University of Minnesota

ABSTRACT

In this appendix I present details of the model and of 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 (technological) 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 present omitted details of the proofs of Propositions 1, 2, and 3 in the paper. In Section 2 I provide an overview of the numerical solution of the model. In Section 3 I discuss in detail model identification. In Section 4 I briefly describe the original U.S. firm dataset of Baker, Gibbs, and Holmström (1994a), on which my work is based. In Section 5 I derive the likelihood function of the model. In Section 6 I present results from a Monte Carlo exercise to show the identifiability of the model's parameters in practice. In Section 7 I derive bounds on the informativeness of jobs at competitors of the firm in my data, based on the estimates of the parameters reported in the paper. Finally, in Section 8 I present estimation results based on a 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: Experimentation; Bandit; Wage Growth; Job Mobility; Turnover
JEL Classification: D22, D83, J24, J31, J44, J62

*I especially thank Hiro Kasahara and Matt Shum for detailed suggestions about how to apply their identification arguments to my framework. I am also grateful to Kyoo il Kim and John Geweke for their many useful comments. Zhen Huo provided outstanding research assistance, and Joan Gieseke and Kathy Rolfe invaluable editorial assistance. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1 Omitted Model Details

1.1 An Example

I consider here 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 my model in the paper, it is sufficient to clarify several key features about equilibrium in my model. First, it makes clear that, even if the model does not include 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 (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.

In this example, I assume that the model has one firm of type (that is, technology) A and at least two firms of type B, so that firm B has a replica. 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 $\alpha_{A2} = \alpha$ and $\beta_{A2} = \beta$), 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 $t = 1, 2$ and features no technological 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 a linear function given by

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

where $\overline{y}_f(\alpha,k) = \alpha_{fk}y_{fHk} + (1-\alpha_{fk})y_{fLk}$, $\overline{y}_f(\beta,k) = \beta_{fk}y_{fHk} + (1-\beta_{fk})y_{fLk}$, $m_{fk} = \overline{y}_f(\beta,k)$, 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 implied specification of the expected output functions in (1).

Trivially, under (2), if the economy starts with each worker's ability known, so that some workers are known to be of low ability ($p = 0$) while others are known to be of high ability ($p = 1$), then the model implies that low ability workers work in job $A1$ and high ability workers work in $B1$. That is, the model implies a rather degenerate distribution of workers to jobs, 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. (And, as the paper documents, the data point to the existence of substantial initial uncertainty about a worker's ability.)

Thus, consider now the more interesting case in which p is interior to $[0, 1]$. From Proposition 3 in the paper, we know that the match surplus value problem for firm A in the first period reduces to

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

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

where $r_{Ak}(p) = \alpha_{Ak}p + \beta_{Ak}(1-p)$ and $r_{B1}(p) = \alpha_{B1}p + \beta_{B1}(1-p)$. The subscript in the value function denotes the time period, the subscript, the firm. I have written the time discounting so that period 2 stands in for a long future. The notation is the same as in the paper with obvious modifications.

Consider solving the model by backward induction from the last period, here, period 2. In the last period, the job assignment decision is static. Clearly, from (2) the static job assignment policy is to assign job $A1$ at low enough priors, assign job $A2$ at intermediate priors, and assign job $B1$ at high enough priors.

More formally, define $\bar{p}_{A2}$ as the static cutoff prior between jobs $A1$ and $A2$, which satisfies $y_A(\bar{p}_{A2}, 1) = y_A(\bar{p}_{A2}, 2)$. Similarly, define $\bar{p}_{B1}$ as the static cutoff prior between jobs $A2$ and $B1$, which satisfies $y_A(\bar{p}_{B1}, 2) = y_B(\bar{p}_{B1}, 1)$. From (1) and (2), it follows that $\bar{p}_{A2} = (c_{A1} - c_{A2})/(m_{A2} - m_{A1})$ and $\bar{p}_{B1} = (c_{A2} - c_{B1})/(m_{B1} - m_{A2})$. Hence, the match surplus value in period 2 is

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

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 $r_{A1}(p) = \alpha_{A1}$, $r_{A2}(p) = \alpha p + \beta(1-p)$, 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 $\hat{p}_{A2}$ at which the firm is indifferent between assigning the worker to jobs $A1$ and $A2$ satisfies

$$\hat{p}_{A2} < \bar{p}_{A2}. \tag{4}$$

Likewise, since job $B1$ has an informational advantage over job $A2$, the cutoff prior $\hat{p}_{B1}$ at which the firm is indifferent between assigning the worker to jobs $A2$ and $B1$ satisfies

$$\hat{p}_{B1} < \bar{p}_{B1}. \tag{5}$$

So a worker with initial prior $p < \hat{p}_{A2}$ starts in job $A1$, a worker with initial prior $p \in [\hat{p}_{A2}, \hat{p}_{B1})$ starts in job $A2$, whereas a worker with initial prior $p \geq \hat{p}_{B1}$ starts in job $B1$.

To say more than this, I need to calculate where such workers are assigned after success and failure in these jobs. For concreteness, I focus on a region of the parameter space in which three conditions hold. First, the worker with the lowest initial prior who is assigned to job $A2$, namely, the worker with prior $\hat{p}_{A2}$, stays in job $A2$ after a success; that is,

$$P_H(\hat{p}_{A2}) < \bar{p}_{B1}. \tag{6}$$

Note also that at $\hat{p}_{A2}$ a worker who fails is demoted to job $A1$, since $P_L(\hat{p}_{A2}) < \hat{p}_{A2}$ if $\alpha > \beta$ and $\hat{p}_{A2} < \bar{p}_{A2}$ by (4). Thus, $P_L(\hat{p}_{A2}) < \bar{p}_{A2}$. Second, the worker with the highest initial prior at job $A2$ is again assigned to job $A2$ after a failure; that is,

$$P_L(\hat{p}_{B1}) \geq \bar{p}_{A2}, \tag{7}$$

which is to be interpreted as $P_L(\hat{p}_{B1} - \varepsilon) \geq \bar{p}_{A2}$ with $\varepsilon > 0$ arbitrarily small. Third, the worker with the lowest initial prior who is assigned to job $B1$, $\hat{p}_{B1}$, is again assigned to $B1$ after a success; that is,

$$P_H(\hat{p}_{B1}) \geq \bar{p}_{B1}, \tag{8}$$

which is also to be interpreted as $P_H(\hat{p}_{B1} - \varepsilon) \geq \bar{p}_{B1}$ with $\varepsilon > 0$ arbitrarily small. Figure 2 illustrates these assumptions graphically.[1]

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

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

The left side of (9) is the value of assigning the worker to job $A1$ in period 1 at prior p. Here I have used the fact that job $A1$ is uninformative about ability, so the prior is not updated after

[1] Observe that the following conditions—$m_{B1} < m_{A2} < m_{A1}$, $m_{A1}+c_{A1} < m_{A2}+c_{A2} < m_{B1}+c_{B1}$, $0 < \hat{p}_{A2} < \hat{p}_{B1} < 1$, $P_H(\hat{p}_{A2}) < \bar{p}_{B1}$, $P_L(\hat{p}_{B1}) > \bar{p}_{A2}$, and $P_H(\hat{p}_{B1}) > \bar{p}_{B1}$—are simultaneously satisfied for the following set of parameters: $\alpha = 0.6$, $\beta = 0.45$, $\delta = 0.1$, $m_{A1} = 3$, $m_{A2} = 2$, $m_{B1} = 0$, $c_{A1} = 1$, $c_{A2} = 5$, and $c_{B1} = 7.5$. Alternatively, these restrictions are satisfied for $\alpha \in [0.5, 0.95]$, $\beta = 0.02$, $\delta = 0.95$, $m_{A1} = 3$, $m_{A2} = 2.9$, $m_{B1} = -10.1645$, $c_{A1} = 1$, $c_{A2} = 1.2$, and $c_{B1} = 14.4145$. By reducing β, the same parameter values would work for values of δ higher than 0.95.

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 $p < \hat{p}_{A2}$. Since $P_{AH1}(p) = P_{AL1}(p) = p < \bar{p}_{A2}$ by (4), the worker is assigned to job $A1$ in period 2 as well.

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 $P_L(\hat{p}_{A2}) < \bar{p}_{A2}$, as argued above.

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

$$(1-\delta)y_A(\hat{p}_{B1}, 2) + \delta\{r_{A2}(\hat{p}_{B1})y_B(P_H(\hat{p}_{B1}), 1) + [1 - r_{A2}(\hat{p}_{B1})]y_A(P_L(\hat{p}_{B1}), 2)\}$$
$$= (1-\delta)y_B(\hat{p}_{B1}, 1) + \delta\left[\hat{p}_{B1}y_B(1,1) + (1-\hat{p}_{B1})y_A(0,1)\right]. \tag{10}$$

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, and by (7) and $P_L(\hat{p}_{B1}) < \hat{p}_{B1} < \bar{p}_{B1}$ by (5), the worker is assigned to job $A2$ after a failure. Under these assumptions, the job assignment policy in the first period is

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

I have set up the example so that the interesting dynamics is generated by workers who start in job $A2$ in the first period. After a success, these workers move from job $A2$ to

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

where I have used the facts that $P_H(\hat{p}_{A2}) < \bar{p}_{B1}$ by (6) and $\bar{p}_{B1} \leq P_H(\hat{p}_{B1})$ by (8). After a failure, these workers move from job $A2$ to

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

where I have used the facts that $P_L(\hat{p}_{A2}) < \bar{p}_{A2}$ by the argument above and $\bar{p}_{A2} \leq P_L(\hat{p}_{B1})$ by (7). 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 signals at different jobs except that I assume that $\alpha_{fk} > \beta_{fk}$ at each job. So job 1 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

$$(0 <)\bar{p}_{A2} < \bar{p}_{B1}(< 1)$$

and the job assignment policy in the second period is

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

Note that the relation between the static cutoffs $\bar{p}_{A2}$ and $\bar{p}_{B1}$ and the dynamic ones $\hat{p}_{A2}$ and $\hat{p}_{B1}$ 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 $\bar{p}_{A2} < \hat{p}_{A2}$ (if we abstract from the trivial case of equality between the two cutoffs). Thus, at $p \in [\bar{p}_{A2}, \hat{p}_{A2})$, even though job $A2$ statically dominates job $A1$, assigning job $A1$ in period 1 is still optimal because the informational advantage of job $A1$ implies that job has a higher (dynamic) match surplus value. In contrast, if job $A2$ is more informative than job $A1$, then the opposite relation holds: $\hat{p}_{A2} < \bar{p}_{A2}$ (again if we abstract 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 $\bar{p}_{B1} < \hat{p}_{B1}$, whereas if job $B1$ is more informative than $A2$, then $\hat{p}_{B1} < \bar{p}_{B1}$.

Note 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. To be concrete, consider job $A2$, which is assigned at all initial priors $p \in [\hat{p}_{A2}, \hat{p}_{B1})$. To indicate what happens after a success, as before I divide this interval into two subintervals, a left subinterval $[\hat{p}_{A2}, P_H^{-1}(\bar{p}_{B1}))$ and a right subinterval $[P_H^{-1}(\bar{p}_{B1}), \hat{p}_{B1})$. In the left subinterval, a success in job $A2$ leads the worker to stay in that job, since $P_H(\hat{p}_{A2}) < \bar{p}_{B1}$ by (6), whereas in the right subinterval a success in job $A2$ leads the worker to move to firm B and work in job $B1$, since $P_H(\hat{p}_{B1}) \geq \bar{p}_{B1}$ by (8).

Likewise, to indicate what happens after a failure in job $A2$, I divide the interval into two other subintervals: a left subinterval $[\hat{p}_{A2}, P_L^{-1}(\bar{p}_{A2}))$ and a right subinterval $[P_L^{-1}(\bar{p}_{A2}), \hat{p}_{B1})$. In the left subinterval, a failure in job $A2$ leads the worker to be demoted to job $A1$, since $P_L(\hat{p}_{A2}) < \bar{p}_{A2}$ by (4), whereas in the right subinterval, a failure in job $A2$ leads the worker to stay in that job, since $P_L(\hat{p}_{B1} - \varepsilon) \geq \bar{p}_{A2}$ by (7).

So far I have discussed what happens to workers who start in job $A2$. A new possibility arises for workers who start in job $A1$, those with initial priors $p \in [0, \hat{p}_{A2})$. To determine job assignment after a success here, I split this interval into three subintervals: a left subinterval $[0, P_H^{-1}(\bar{p}_{A2}))$, a middle subinterval $[P_H^{-1}(\bar{p}_{A2}), P_H^{-1}(\bar{p}_{B1}))$, and a right subinterval $[P_H^{-1}(\bar{p}_{B1}), \hat{p}_{A2})$. In the left subinterval, a success leads the worker to stay in job $A1$, in the middle subinterval, success leads to a move to job $A2$, and in the right subinterval, success leads the worker to move to firm B's job $B1$. Of course, for this to happen, job $A1$ has to be sufficiently informative.

Likewise, workers who start in job $B1$ have three possibilities after a failure: those in a left subinterval move to job $A1$, those in a middle subinterval move to $A2$, and those in a right subinterval stay in $B1$.

So far I have assumed firm A has two jobs and firm B has one job. The more general case—with three jobs at firm A and three jobs at each of the other firms B, C, D, and so on, each of which has a replica—yields to many more cases. In this sense, even this very simple model can generate rich patterns of job mobility. When I incorporate into this model technological human capital acquisition, productivity shocks, and separation shocks, the model is flexible enough to generate the rich non-linear, non-monotone patterns of mobility I observe in the data.

1.2 Omitted Proofs

Proof of Proposition 1. Consider first equilibrium states at which firm A employs the worker. Observe that when firm A employs the worker, the match surplus value between firm A and the worker is given by

$$V^A(s_t, \boldsymbol{\varepsilon}_t) = \max_{k \in K^A} \left\{ (1-\delta)[y_A(s_t,k) + \varepsilon_{Akt}] + \delta(1-\eta_{Akt}) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k) dF(\boldsymbol{\varepsilon}_{t+1}) \right\} \tag{13}$$

by the argument in the paper, which can be rewritten as

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

with $k_{At} = k_A(s_t, \boldsymbol{\varepsilon}_t)$. Compute now the match surplus value between firm A and the worker if, instead of accepting firm A's offer, the worker accepts firm f's offer. Based on equilibrium strategies, the match surplus value in such a case would equal

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

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

by definition of $V^A(\cdot)$. Now, firm's optimality and the worker's indifference between firm A's and firm f's offers in any Markov perfect equilibrium (MPE), at states at which firm A employs the worker, imply that

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

$$\geq (1-\delta)w_f(s_t, \boldsymbol{\varepsilon}_t) + \delta(1-\eta_{fk_{ft}t}) \int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1}, \boldsymbol{\varepsilon}_{t+1}|s_t,k_{ft}) dF(\boldsymbol{\varepsilon}_{t+1}). \tag{14}$$

Consider now equilibrium states at which firm f employs the worker. Now the match surplus value between firm A and the worker in this case is given by

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

The match surplus value between firm A and the worker—if, instead of accepting firm f's offer, the worker accepts firm A's offer—is given by

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

where, by the cautious equilibrium restriction, the choice of $k_{At}=k_A(s_t,\boldsymbol{\varepsilon}_t)$ satisfies (13). At equilibrium states at which the worker accepts firm f's offer, the worker must weakly prefer firm f's offer over firm A's offer, whereas firm A must weakly prefer not employing over employing the worker. (The worker is indifferent between the two offers when firm A happens to be the second-best firm; firm A is indifferent between employing and not employing the worker when it happens to be the second-best firm and the MPE is cautious.) Hence,

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

$$\geq (1-\delta)[y_A(s_t,k_{At})+\varepsilon_{Ak_{At}t}]+\delta(1-\eta_{Ak_{At}t})\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})dF(\boldsymbol{\varepsilon}_{t+1}). \quad (15)$$

By combining (14) at states at which firm A employs the worker and (15) at states at which firm f employs the worker, we see that $V^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1})$ equals

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

$$\left.(1-\delta)w_f(s_t,\boldsymbol{\varepsilon}_t)+\delta(1-\eta_{fk_{ft}t})\int_{\boldsymbol{\varepsilon}_{t+1}} EV^A(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{ft})dF(\boldsymbol{\varepsilon}_{t+1})\right\}, \quad (16)$$

which proves Proposition 1. ∎

Proof of Proposition 2. The worker's indifference between firm A's and firm f's offers implies that

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

$$=(1-\delta)w_f(s_t,\boldsymbol{\varepsilon}_t)+\delta(1-\eta_{fk_{ft}t})\int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{ft})dF(\boldsymbol{\varepsilon}_{t+1}). \quad (17)$$

Now, by rearranging the cautious equilibrium restriction for firm f, we see that

$$w_f(s_t,\boldsymbol{\varepsilon}_t)=y_f(s_t,k_{ft})+\varepsilon_{fk_{ft}t}+\frac{\delta}{1-\delta}\int_{\boldsymbol{\varepsilon}_{t+1}}\Big[(1-\eta_{fk_{ft}t})E\Pi^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{ft})$$

$$-(1-\eta_{Ak_{At}t})E\Pi^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})\Big]\,dF(\boldsymbol{\varepsilon}_{t+1}). \quad (18)$$

Substituting this last expression into (17) yields that

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

$$-\delta(1-\eta_{fk_{ft}t})\int_{\boldsymbol{\varepsilon}_{t+1}} EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{ft})dF(\boldsymbol{\varepsilon}_{t+1})=(1-\delta)[y_f(s_t,k_{ft})+\varepsilon_{fk_{ft}t}]$$

$$+\delta\left[\int_{\boldsymbol{\varepsilon}_{t+1}}(1-\eta_{fk_{ft}t})E\Pi^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{ft})-(1-\eta_{Ak_{At}t})E\Pi^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})\right]dF(\boldsymbol{\varepsilon}_{t+1})$$

or, equivalently,

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

$$+\frac{\delta}{1-\delta}\int_{\boldsymbol{\varepsilon}_{t+1}}\Big\{(1-\eta_{fk_{ft}t})[E\Pi^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{ft})+EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{ft})]$$

$$-(1-\eta_{Ak_{At}t})\Big[E\Pi^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})+EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|s_t,k_{At})\Big]\Big\}\,dF(\boldsymbol{\varepsilon}_{t+1}).$$

Since, by definition, $EV^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|\cdot)=E\Pi^f(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|\cdot)+EV^w(s_{t+1},\boldsymbol{\varepsilon}_{t+1}|\cdot)$, Proposition 2 follows. ∎

Proof of Proposition 3. From (18), we see that $w_f(s_t,\boldsymbol{\varepsilon}_t)=y_f(s_t,k_{ft})+\varepsilon_{fk_{ft}t}$ when the difference between the relevant continuation values is zero. This, together with (16), proves the desired result. ∎

2 Numerical Solution of the Model

My numerical approach to computing the match surplus value and the job-specific match surplus values builds on the work of Rust (1987, 1988, 1994) on the solution and estimation of stochastic dynamic discrete choice programming problems. Here I show how I apply Rust's method to the equilibrium best-response retention and job assignment problem that each firm faces in my model.

2.1 Decision Problem

I first describe the match surplus problem of my firm, *firm A*. I make 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, so that the continuation value at tenure 7 is the stationary value at tenure 8. Thus, my match surplus maximization problem consists of both non-stationary and stationary parts. Nonetheless, adapting the work of Rust (1994, p. 3108) to my problem is straightforward.

To render the match surplus problem stationary from tenure 8 on, I make two assumptions. First, I assume that from tenure $t=8$ on, the stock of technological human capital acquired by any 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 retained 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 tenures at any job k by η_k.

Consider now the stationary match surplus problem from the eighth year of tenure on. Omit the firm superscript and the firm and tenure subscripts and denote by $\boldsymbol{\varepsilon}$ the current value of productivity shocks and by $\boldsymbol{\varepsilon}'$ their future value. Then the value of firm A's problem is

$$V_8(p_i, \boldsymbol{\varepsilon}) = \max_{k \in \{0,1,2,3\}} \{V_8(p_i, \boldsymbol{\varepsilon}, k)\} = \max_{k \in \{0,1,2,3\}} \{v_8(p_i, k) + \varepsilon_k\}, \tag{19}$$

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

$$v_8(p_i, k) = (1-\delta)y(p_i, k) + \delta(1-\eta_k)r_k(p_i) \int_{\varepsilon'} V_8(P_{Hk}(p_i), \boldsymbol{\varepsilon}')dF(\boldsymbol{\varepsilon}')$$
$$+\delta(1-\eta_k)[1-r_k(p_i)] \int_{\varepsilon'} V_8(P_{Lk}(p_i), \boldsymbol{\varepsilon}')dF(\boldsymbol{\varepsilon}') \tag{20}$$

and the value of separation, $v_8(p_i, 0)$, is approximated as discussed in the paper.

For tenures t ranging from 1 through 7, the match surplus problem has state $s_{it} = (p_{it}, t-1, k_{t-1})$ and value

$$V(p_{it}, t-1, k_{t-1}, \boldsymbol{\varepsilon}_t) = \max_{k \in \{0,1,2,3\}} \{V(p_{it}, t-1, k_{t-1}, \boldsymbol{\varepsilon}_t, k)\} = \max_{k \in \{0,1,2,3\}} \{v(p_{it}, t-1, k_{t-1}, k) + \varepsilon_{kt}\}, \tag{21}$$

where

$$v(p_{it}, t-1, k_{t-1}, k) = (1-\delta)y(p_{it}, t-1, k_{t-1}, k) + \delta(1-\eta_{kt})r_k(p_{it}) \int_{\boldsymbol{\varepsilon}_{t+1}} V(P_{Hk}(p_{it}), t, k, \boldsymbol{\varepsilon}_{t+1})dF(\boldsymbol{\varepsilon}_{t+1})$$
$$+\delta(1-\eta_{kt})[1-r_k(p_{it})] \int_{\boldsymbol{\varepsilon}_{t+1}} V(P_{Lk}(p_{it}), t, k, \boldsymbol{\varepsilon}_{t+1})dF(\boldsymbol{\varepsilon}_{t+1}) \tag{22}$$

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

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

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

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 I assume in the paper, the location parameter of the distribution of each shock, a, must equal $-\gamma$ and the variance parameter, b, must equal 1.

Observe that, under this distributional assumption, at any tenure the match surplus problem is akin to a standard dynamic multinomial logit problem. Hence, standard techniques can be applied to derive probabilities of observed job assignment and separation.

I solve for the probability of an observed assignment in three steps, as follows. In the first step I solve for firm A's match surplus value function at tenure $t \geq 8$, and in the second step I use this computed value as the terminal value in a backward induction routine that solves for firm A's match surplus value from tenure $t = 1$ until $t = 7$. In the third step I derive the probabilities of interest.

First I compute the value function for firm A's problem from tenure $t = 8$. Note that at any prior p'_i,

$$\int_{\boldsymbol{\varepsilon}'} V_8(p'_i, \boldsymbol{\varepsilon}')dF(\boldsymbol{\varepsilon}') = \int_{\boldsymbol{\varepsilon}'} \max_{k' \in \{0,1,2,3\}} \{v_8(p'_i, k') + \varepsilon'_{k'}\}dF(\boldsymbol{\varepsilon}') = \log\left\{\sum_{k' \in \{0,1,2,3\}} \exp\left[v_8(p'_i, k')\right]\right\}, \quad (23)$$

which implies that $v_8(p_i, k)$ from (20) can be rewritten as

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

$$+\delta(1-\eta_k)[1-r_k(p_i)]\log\left\{\sum_{k' \in \{0,1,2,3\}} \exp\left[v_8(P_{Lk}(p_i), k')\right]\right\}. \quad (24)$$

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

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

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

$k = 1, 2, 3$, where Γ is an operator on the function $v_8(\cdot)$. Note that here I follow the formulation in Rust (1988, 1994), where 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 is to use the numerical solution of the match surplus value at $t \geq 8$ as input to the backward induction recursion defining the match surplus values at the remaining tenure dates. Specifically, consider tenures between $t = 2$ and $t = 7$. For these tenures, given (21)–(24), I can compute $v(p_{it}, t-1, k_{t-1}, k)$ as

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

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

where the continuation value at $t = 7$ is $v(p_{it+1}, t, k, k') = v_8(p_{it+1}, k')$. Next consider the match surplus value functions at $t = 1$. These value functions differ from the values just derived because, as argued in the paper, the state only consists of p_{i1} rather than p_{it}, $t-1$, and k_{t-1}.[2]

For the third step, I compute the probability of an observed assignment by taking as input the match surplus values at each tenure calculated above. From Rust (1994), then, the probability of an observed assignment $k_t = k$ for a manager of type i at any tenure between $t = 2$ and $t = 7$ is given by

$$\Pr(k_t = k | p_{it}, t-1, k_{t-1}) = (1-\eta_{k_{t-1}t-1}) \frac{\exp\{v(p_{it}, t-1, k_{t-1}, k)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v(p_{it}, t-1, k_{t-1}, k')\}} \tag{25}$$

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

$$\Pr(k_t = 0 | p_{it}, t-1, k_{t-1}) = (1-\eta_{k_{t-1}t-1}) \frac{\exp\{v(p_{it}, t-1, k_{t-1}, 0)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v(p_{it}, t-1, k_{t-1}, k')\}} + \eta_{k_{t-1}t-1} \tag{26}$$

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

$$\Pr(k_1 = k | p_{i1}) = \frac{\exp\{v(p_{i1}, k)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v(p_{i1}, k')\}} \tag{27}$$

whereas the probability of an assignment k with $1 \leq k \leq 3$ at $t = 8$ is given by

$$\Pr(k_t = k | p_{it}) = (1-\eta_{k_{t-1}t-1}) \frac{\exp\{v_8(p_{it}, k)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v_8(p_{it}, k')\}}, \tag{28}$$

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

$$\Pr(k_t = 0 | p_{it}) = (1-\eta_{k_{t-1}t-1}) \frac{\exp\{v_8(p_{it}, 0)\}}{\sum_{k' \in \{0,1,2,3\}} \exp\{v_8(p_{it}, k')\}} + \eta_{k_{t-1}t-1}. \tag{29}$$

2.3 Prior Grid

I compute recursively the job-specific match surplus values $v_8(p_i, k)$ by value function iteration. I discretize the support of p_i, $[0, 1]$, to a uniform grid of 100 equidistant points. (I also experimented with

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

finer grids, but results are virtually unaffected, and the increase in computational cost is substantial.)

Observe that under 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). Such a model, in fact, would have only one α and one β for all jobs, because all 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 belief reached after the experience of a success and a failure, or after the experience of a failure and a success, is $P_L(P_H(p_{it})) = P_H(P_L(p_{it})) = p_{it}$.

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

I view my assumptions as allowing for a more flexible specification for the belief process than is typical in the literature. Conversely, information on job assignment, performance ratings, and wages for managers experiencing different sequences of performance ratings helps pin down α_k and β_k at the different jobs.

3 Identification

As discussed below and in the paper, in estimation I impose a number of restrictions, which lead to 75 parameters to be estimated for the sample of entrants into the firm at Level 1. The model has a sizeable number of parameters, but it is also being fit to rich data on the sequence of yearly job assignments, paid wages, and recorded performance ratings for more than 1,400 managers over eight years. I now discuss identification of the model based on these data. The Monte Carlo results below provide additional evidence on the fact that the variation in the data helps pin down quite precisely the model's parameters.

3.1 Discrete-Choice Component of the Model: Retention and Job Assignment

Consider the retention 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. Recall that beliefs, which are unobserved to the econometrician, are modeled as a nonparametric finite mixture distribution with known components. In the model, job assignment depends on beliefs, a manager's human capital, and the realization of (type I extreme value) productivity shocks. Hence, the discrete choice component of my model features a nonparametric mixture of parametric component distributions.

Hence, the identification of the discrete choice component of my model amounts to the identification of the process $\{k_t, p_t, h_t\}$ 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

$$\Pr(k_{t+1}, p_{t+1}, h_{t+1}|p_t, h_t, k_t) = \Pr(k_{t+1}|p_t, h_t, k_t, p_{t+1}, h_{t+1}) \Pr(h_{t+1}|p_t, h_t, k_t, p_{t+1}) \Pr(p_{t+1}|p_t, h_t, k_t), \tag{30}$$

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

$$\Pr(k_{t+1}, p_{t+1}, h_{t+1}|p_t, h_t, k_t) = \Pr(k_{t+1}|p_{t+1}, h_{t+1}) \Pr(h_{t+1}|h_t, k_t) \Pr(p_{t+1}|p_t, k_t). \tag{31}$$

The equality of the first term on the right side of (30) with the first term on the right side of (31) follows because the next period job assignment depends only on the next period prior and human capital by the Markovian nature of the match surplus problem and by Rust's formulation, as discussed above. The equality of the second terms in these expressions follows because the next period human capital depends only on current human capital and job assignment by assumption (see in the paper the specification of the process of technological human capital acquisition). The equality of the third terms follows because the next period prior depends only on the current prior and job assignment by Bayes' rule.

As for the productivity shocks determining $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$, note that the role of productivity shocks is auxiliary to the main focus of the estimation exercise, which is the recovery of the primitive parameters of the informational and technological human capital process and of the firm's technology. Productivity shocks simply contribute to make the job choice of the employing firm 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 next period to different jobs after having been assigned to the *same* job and experiencing the *same* recorded performance rating in the current period.

Here I first provide an intuitive argument for the parametric identification of the model based on specific moments of the distribution of the observables. Then, I provide an argument for the nonparametric identification of the process governing the evolution of the state variables, p_t and h_t, and the choice variable, k_t, based on Hu and Shum (2012) and Kasahara and Shimotsu (2009).

The reason for providing these two arguments is as follows. The logic of the arguments of Hu and Shum (2012) and Kasahara and Shimotsu (2009) for the nonparametric identification of dynamic discrete choice models suggest 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. However, Hu and Shum (2012) and Kasahara and Shimotsu (2009) provide arguments for identification that are based on high-level assumptions, and neither exploits restrictions on outcomes implied by the underlying economic model.

Therefore, here I start by providing a moment-based argument for identification that is transparent and based on the implications of the model about level assignments, performance ratings, and wages, which I discuss in the paper when presenting the data and the main descriptive statistics. This moment-based argument is based on the logic that different moments of the distribution of observed job assignments, performance ratings, and wages naturally identify different structural parameters in light of the restrictions implied by the model.

3.1.1 A Moment-Based Argument

I start with a simple example illustrating how the combination of assumptions and functional form restrictions of the theory provide a source of identification of the mixture discrete choice component of my model. I then present the moment-based argument for the general case of my model.

An Illustrative Example: Parametric Local Identification. Here I provide an argument for the local identification of a semiparametric mixture model of discrete choice with type I extreme value components and fixed (two, for simplicity) number of components. The identification of the more general case with multiple components follows the same logic. 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 $F(\cdot)$ and their probability density function by $f(\cdot)$. This model implies

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

$$= \frac{q}{1 + \exp\{-\rho(z) - u_1\}} + \frac{(1 - q)}{1 + \exp\{-\rho(z) - u_2\}} = qF(\rho(z) + u_1) + (1 - q)F(\rho(z) + u_2). \quad (32)$$

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

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

and

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

Note for later that

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

Proposition 1. *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. Specifically, in the following I will show the positive definiteness of the information matrix for model (32). 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

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

$$= \frac{y\,[1-P(z)] - (1-y)P(z)}{P(z)\,[1-P(z)]}\frac{\partial P(z)}{\partial\vartheta} = \frac{y-P(z)}{P(z)\,[1-P(z)]}\frac{\partial P(z)}{\partial\vartheta}.$$

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

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

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

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

$$\begin{aligned}
\frac{\partial P(z)}{\partial q} &= F(\rho(z)+u_1) - F(\rho(z)+u_2) = 0 \\
\frac{\partial P(z)}{\partial u_1} &= qf(\rho(z)+u_1) = 0 \\
\frac{\partial P(z)}{\partial u_2} &= (1-q)f(\rho(z)+u_2) = 0.
\end{aligned}$$

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

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

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

$$w_1[f(\rho(z)+u_1) - f(\rho(z)+u_2)]\rho'(z) + w_2 q f'(\rho(z)+u_1)\rho'(z)$$

$$+w_3(1-q)f'(\rho(z)+u_2)\rho'(z) = 0. \tag{33}$$

Let $r(z) = f(\rho(z)+u_1)/f(\rho(z)+u_2)$. By dividing the left-hand side and the right-hand side of (33) by $f(\rho(z)+u_2)$ and $\rho'(z)$, I obtain

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

It follows

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

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

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

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

or, equivalently,

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

$$-\underbrace{[w_1 + w_3(1-q)\left(1-\exp\{-\rho(z)-u_2\}\right)F(\rho(z)+u_2)]}_{B} = 0. \tag{34}$$

Since $r(z)$ is a non-constant exponential function of z, (34) holds for all $z \in Z^*$ only if both terms A and B in (34) are zero for each $z \in Z^*$, that is, if

$$w_1 - w_2 q\frac{1-\exp\{-\rho(z)-u_1\}}{1+\exp\{-\rho(z)-u_1\}} = 0 \tag{35}$$

and

$$w_1 + w_3(1-q)\frac{1-\exp\{-\rho(z)-u_2\}}{1+\exp\{-\rho(z)-u_2\}} = 0. \tag{36}$$

Now a necessary condition for (35) and (36) 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 (35) with respect to z, evaluated at $z \in Z^*$, it follows

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

which, since $\rho'(z)$ is different from zero by assumption and $(1 + \exp\{-\rho(z) - u_1\})^2$ is also different from zero, can be rewritten as

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

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

or, equivalently,

$$2 w_2 q \exp\{-\rho(z) - u_1\} = 0.$$

Since $q \in (0, 1)$, it follows $w_2 = 0$. Hence, by (35) it also follows that $w_1 = 0$.

Similarly, taking the derivative of (36) with respect to z, evaluated at $z \in Z^*$, it follows

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

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

which, since $\rho'(z)$ is different from zero by assumption and $(1 + \exp\{-\rho(z) - u_2\})^2$ 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,

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

Since $q \in (0, 1)$, it follows $w_3 = 0$. ∎

Note that in principle the same 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. The reason is that if $\rho(z)$ stands in for the relevant value function in a dynamic version of this problem à la Rust (1987), then $\rho(z)$, u_1, and u_2, for instance, are no longer linearly separable. For this reason, I now turn to a more general moment-based argument.

A Moment-Based Argument. First, note that the proportions of managers entering the firm between 1970 and 1979, who are assigned to Levels 1, 2, and 3 between tenure $t = 1$ and $t = 7$, provide a set of moments that identify the output parameters a_{kt}, $b_{kk_{t-1}t}$, and c_{kt}, given the prior beliefs $\{p_{i1}\}_{i=1}^4$, the parameters $\{\alpha_k, \beta_k\}_{k=1}^3$, and the discount factor δ (which is fixed at 0.95 in estimation). Next, note that the only output parameter to be identified in the eighth tenure year is c_{38}. To see this, recall that, as discussed in the Appendix in the paper, the exogenous separation rates at $t \geq 8$ at Level 3 are normalized at zero. Then, c_{38} can easily be recovered from the empirical frequency of separations in $t = 8$ in the sample. Similarly, exogenous separation rate parameters are identified by the tenure profile of the hazard rate of separation at each level. Note that in my framework exogenous separations differ from endogenous separations in that the incidence of exogenous separations does not vary with beliefs or 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). This feature of the data suggests that exogenous separations are a quantitatively important determinant of turnover, as estimation results confirm.

Next observe that after entry, the same distribution of managers across Levels 1, 2, and 3 can be generated through different patterns of transitions of managers across levels. Hence, given δ, at each level and tenure the hazard rates of retention at a level (no job change), promotion from a level to a higher one (positive job change), and demotion from a level to a lower one (negative job change) provide information on $\{p_{i1}\}_{i=1}^{4}$ and $\{\alpha_k, \beta_k\}_{k=1}^{3}$. Specifically, the extent to which job switching varies with the number of times managers have been assigned to a given level helps identify $\{\alpha_k, \beta_k\}_{k=1}^{3}$, whereas differences in the assignment probabilities early versus late in tenure help identify the initial prior distribution. (See Crawford and Shum (2005) for a similar argument.) Also, recall the discussion in the paper of the descriptive statistics from the sample. The timing of promotions and the fraction of promoted managers at different tenures, as well as the distribution of ratings among promoted and unpromoted managers, help identify the number and fraction of managers of different skill types and, thus, the distribution of initial priors. As explained below, the observed distributions of wages at each level and tenure provide further moments that help pin down beliefs.

Lastly, 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. Indeed, note that differently from most estimation exercises of dynamic learning models, my data include information about all managers' performance ratings in each year of employment, which, under the model, correspond to the distribution of output signals. To see how this information provides a crucial source of identification for $\{\alpha_k, \beta_k\}_{k=1}^{3}$, note that the probability of a high rating for manager n of skill type i at tenure t is

$$\Pr(R^o_{int} = 1 | L^o_{int}, t) = \sum_{\theta \in \{\alpha, \beta\}} \Pr(\theta | i, e_n, t) [\Pr(R^o_{int} = 1 | R_{int} = 1, L^o_{int}, t) \Pr(R_{int} = 1 | L^o_{int}, \theta)$$

$$+ \Pr(R^o_{int} = 1 | R_{int} = 0, L^o_{int}, t) \Pr(R_{int} = 0 | L^o_{int}, \theta)],$$

where $p_{i1} = \Pr(\alpha | i, e_n, 1)$, $e_n = (age_n, edu_n, year_n)$ denotes the vector of observed characteristics of manager n, namely, age at entry, education at entry, and calendar year of entry, R^o_{int} denotes the observed rating, L^o_{int} denotes the observed level, and R_{int} denotes the realized rating (of manager n of type i in period t). Equivalently,

$$\Pr(R^o_{int} = 1 | L^o_{int}, t) = \Pr(R^o_{int} = 1 | R_{int} = 1, L^o_{int}, t) \sum_{\theta \in \{\alpha, \beta\}} \Pr(\theta | i, e_n, t) \Pr(R_{int} = 1 | L^o_{int}, \theta)$$

$$+ \Pr(R^o_{int} = 1 | R_{int} = 0, L^o_{int}, t) \sum_{\theta \in \{\alpha, \beta\}} \Pr(\theta | i, e_n, t) [1 - \Pr(R_{int} = 1 | L^o_{int}, \theta)],$$

which can be rewritten as

$$\Pr(R^o_{int}=1|L^o_{int},t)=E_0(L^o_{int},t)+[1-E_0(L^o_{int},t)-E_1(L^o_{int},t)]\sum_{\theta\in\{\alpha,\beta\}}\Pr(\theta|i,e_n,t)\Pr(R_{int}=1|L^o_{int},\theta) \tag{37}$$

with $E_0(L^o_{int},t)=\Pr(R^o_{int}=1|R_{int}=0,L^o_{int},t)$ and $E_1(L^o_{int},t)=\Pr(R^o_{int}=0|R_{int}=1,L^o_{int},t)$.

I now explain how each term in (37) is identified. Consider first the classification error rates, $E_0(L^o_{int},t)$ and $E_1(L^o_{int},t)$. Lewbel (2000) proves that binary choice models in which classification error is covariate-dependent are semiparametrically identified. In particular, Lewbel (2000) shows that classification error rates are nonparametrically identified under the assumption that a certain monotonicity condition, already invoked by Hausman, Abrevaya, and Scott-Morton (1998), is satisfied. In my setting, this monotonicity condition corresponds to the requirement that the probability of a reported high rating increases with the probability of a true high rating, and this condition is equivalent to the restriction that $d_1>0$, which is satisfied by my estimates.

Consider the last term in (37). For given $\{p_{i1}\}_{i=1}^4$, which is identified as discussed, the distribution of performance ratings among managers continuously assigned to Level 1 identifies α_1 and β_1. In turn, given $\{p_{i1}\}_{i=1}^4$, α_1, and β_1, the distribution of ratings among managers in their first year at Level 2, after promotion from Level 1 to Level 2, identifies α_2 and β_2. Hence, $\{\alpha_k,\beta_k\}_{k=1}^2$ are identified.

A similar argument shows that α_3 and β_3 are also identified by the distribution of performance ratings of managers in their first year at Level 3, after having been assigned to Levels 1 and 2. However, as explained in the paper, in estimation I do not use information on performance ratings of managers at Level 3 but, rather, rely on the hazard rate of retention at Level 3 to pin down α_3 and β_3, as detailed above.[3]

3.1.2 An Argument Based on Hu and Shum (2012)

I divide the discussion into three parts. First, I show that my problem can be cast into the framework of Hu and Shum (2012). Second, I discuss the identification of the processes for the two state variables, $\Pr(p_{t+1}|p_t,k_t)$ and $\Pr(h_{t+1}|h_t,k_t)$. Third, I discuss the identification of the process for the choice variable, $\Pr(k_{t+1}|p_{t+1},h_{t+1})$, and the primitive parameters determining $\Pr(k_{t+1}|p_{t+1},h_{t+1})$ according to (25)–(29).

Reformulation à la Hu and Shum (2012). To see how my identification problem can be cast 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

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

[3] Note also that 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.

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

that is, when $\Pr(Y_t|\cdot)$ and $\Pr(M_t|\cdot)$ do not depend on X_{t-1}^*. My problem is an instance of theirs in that

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

To see this, 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 (31) it follows that

$$\begin{aligned} \Pr(Y_t, M_t, X_t^*|Y_{t-1}, M_{t-1}, X_{t-1}^*) &= \Pr(k_t, h_t, p_t|k_{t-1}, h_{t-1}, p_{t-1}) \\ &= \Pr(k_t|h_t, p_t) \Pr(h_t|k_{t-1}, h_{t-1}) \Pr(p_t|k_{t-1}, p_{t-1}) \\ &= \Pr(Y_t|M_t, X_t^*) \Pr(M_t|Y_{t-1}, M_{t-1}) \Pr(X_t^*|Y_{t-1}, X_{t-1}^*). \end{aligned} \tag{38}$$

Processes for the State Variables. Here I provide an argument for the identification of the processes for p_t, which represents the informational human capital of a manager, and for h_t, which represents the technological human capital of a manager. In this argument, I initially treat the distribution of signals, governed by $\{\alpha_k, \beta_k\}_{k=1}^3$, as known.

If the distribution of signals is known, then the law of motion for the state is known up to the initial prior distribution. To see why, consider first $\Pr(h_{t+1}|h_t, k_t)$. Note that conditional on current human capital h_t and job assignment k_t, the law of motion for human capital is deterministic, in that $\Pr(h_{t+1}|h_t, k_t) = \Pr(t, k_t|t-1, k_{t-1}, k_t)$.

Consider now $\Pr(p_{t+1}|p_t, k_t)$. Observe that $\Pr(p_{t+1}|p_t, k_t)$ 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 obviously implied by the model, because θ is unknown to all model agents. That the law of motion $\Pr(p_{t+1}|p_t, k_t)$ does not depend on a manager's skill type i follows because the probabilities $\{\alpha_k, \beta_k\}_{k=1}^3$ governing the output signals about ability 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 job assignment conditional on p_t. Hence, if the distribution of output signals is known, then the only remaining unknown object is the distribution of initial conditions $\Pr(p_1|p_0, k_0)$, which, as discussed in the paper, is assumed to reduce to $\Pr(p_1)$. Thus, the identification problem for the law of motion for the state reduces to identifying the distribution of initial priors, $\Pr(p_1)$.

To see how $\Pr(p_1)$ is identified, observe that the panel dimension of the sample ($T \geq 5$) implies that the identification result by Hu and Shum (2012) for dynamic discrete choice models applies. Hu and Shum (2012) consider a general class of dynamic discrete choice problems with serially correlated, 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 all nonparametrically identified. In particular, their result covers frameworks like mine in which the unobserved state variable (that is, the prior) is time-varying and can evolve depending on past values of the observed state and choice variables. Their result applied to my model then ensures the nonparametric identification of $\Pr(p_1)$.[4]

Now consider the case in which the distribution of signals is not known. As just mentioned, the law of motion $\Pr(p_{t+1}|p_t, k_t)$ for the unobserved state p_t is nonparametrically identified, according to the result of Hu and Shum (2012). Next, notice that Bayesian updating provides the functional form for the state dependence of the belief process. So in my framework the result of Hu and Shum (2012) can be specifically invoked just to establish the nonparametric identification of the distribution of observed assignments (that is, the choice probabilities), output signals, and the initial prior.

Process for the Choice Variable. I now turn to the argument for the nonparametric identification of the process for the choice variable, namely, the probabilities of job assignment and (endogenous) separation $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$, as well as the output parameters, a_{kt}, $b_{kk_{t-1}t}$, and c_{kt}, determining it.

As mentioned, the identification result of Hu and Shum (2012) implies that the conditional choice probabilities, here, $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$, the law of motion for the unobserved state, here, $\Pr(p_{t+1}|p_t, k_t)$, and the distribution of initial conditions, here, $\Pr(p_1)$, are all nonparametrically identified. So $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$ is identified. Hence, what is left to argue is that, given $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$, output parameters are also identified.

Consider the firm's problem in the first seven years of tenure, taking as given the value function in the eighth year of tenure. For these initial years, I can identify the output parameters based on the result of Magnac and Thesmar (2002) on the nonparametric identification of models of dynamic discrete choice. Specifically, first note that my formulation implies that the law of motion for beliefs, human capital, and productivity shocks satisfies the usual conditional independence assumption common in models of dynamic discrete choice. That is, 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. Also, as discussed in the paper, I treat a manager's employment at the second-best competitor of my firm as the reference alternative and set its value to zero. Hence, the result of Magnac and Thesmar (2002) (in the version without unobserved fixed effects) ensures that per-period utilities are nonparametrically identified.

Consider now the eighth period of tenure. In this last period, the firm solves an infinite-horizon match surplus maximization problem in which, however, only one parameter of static expected utility is unknown. Thus, if all other parameters are identified, we can easily see that this parameter is too. Combining these two arguments, I conclude that output parameters are identified.

So far I have relied on Hu and Shum (2012) to argue identification as my model falls into their basic framework. However, when the unobserved state variable is continuous, as in my model, their nonparametric identification result 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).) Therefore, I find it useful to supplement their argument for identification with a more direct and constructive argument based on Kasahara

[4] I specify $\Pr(p_1)$ as a finite mixture with known components, $i = 1, \ldots, I$. Notice the usual lack of identification of the prior distribution 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.

and Shimotsu (2009).

3.1.3 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. However, their results do not apply when choice probabilities are simultaneously nonstationary and state-dependent. Even more critically for my application, they consider models in which the unobserved state is time-invariant. Hence, their results do not immediately apply to frameworks like mine in which the unobserved state, here the prior, evolves over time.

Suppose, however, that the distribution of signals is known. (See, for instance, the logic following (37).) Recall, as mentioned, that Bayesian updating implies that the state-dependent process for beliefs is known up to the initial prior and the distribution of realized performance. Then, the identification of the discrete choice component of the model reduces to the identification of the nonparametric mixture of initial priors, that is, the unobserved distribution of types, and of the type-specific components, that is, the job assignment probabilities conditional on the type-specific initial prior. (Recall the argument above showing that the (conditional) process of technological human capital acquisition is completely determined by the process of beliefs and job assignment.) Hence, by applying the argument in Kasahara and Shimotsu with $\{\alpha_k, \beta_k\}_{k=1}^3$ known, I can conclude that the distribution of the prior, $\Pr(p_1)$, and the process $\Pr(k_{t+1}|p_{t+1}, h_{t+1})$ 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 these 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, interpreted as a case in which the model admits only time-invariant covariates, the time-series variation in observed choices substitutes for the required time variation of covariates. Specifically, the sufficient conditions for identification reduce to the following: (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).

3.2 Continuous-Choice Component of the Model: Wages

Consider now the identification of the wage parameters. Recall 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_n^2, and edu_n, are equal at Level 1 and Level 2, and denote their common value by ϖ_1, ϖ_2, and ϖ_3, whereas I denote by ϖ_{13}, ϖ_{23}, and ϖ_{33} the coefficients on age_n, age_n^2, and edu_n at Level 3. I also restrict the coefficients on the dummies for the year of entry so that $\varpi_{ym} = 0$ for $m = 0, 1, 2, 3$, and $\varpi_{y4} = \varpi_{y5}$.

As for the remaining parameters, recall from the Appendix in the paper that I allow for a tenure trend only at Level 1, parameterized as $\omega_{11t} = \omega_{111} I(t < 5) + \omega_{115} I(t \geq 5)$ with $\omega_{115} = -\omega_{111}$, to account for the progressively greater proportion of managers at Level 1, who are paid low wages from the fifth tenure year on. Recall that, to conserve on parameters, I assume that ω_{2ik}, the coefficient on the first-degree prior term, does not vary with k. I also assume that the coefficient on the second-degree prior term in the polynomial for the learning and human capital premium in wages is identical across levels and denote its common value across levels by ω_3. Lastly, recall that I assume that the variance of the lognormal disturbance u_{inkt} does not vary across skill types at Level 3. Hence, the estimated wage parameters are $\{\varpi_{0i1}\}_{i=1}^{4}$, $(\varpi_1, \varpi_2, \varpi_3, \varpi_{13}, \varpi_{23}, \varpi_{33})$, $\{\varpi_{yd}\}_{d=5}^{9}$, ω_{111}, and $\{\omega_{2i}\}_{i=1}^{4}$, whereas the estimated variance parameters are $(\{\sigma_{1i}, \sigma_{2i}\}_{i=1}^{4}, \sigma_3)$. (See the Appendix in the paper for a discussion of specific parameter restrictions.)

To see how these parameters are identified, note from the expressions in the paper for the estimated wage equation that conditional on beliefs, (log) wages can be formally thought of as determined by a linear regression model with nonparametric random intercept and slope, in which the term $\omega_{ik}(age_n, edu_n, year_n)$, parameterized by ϖ_{0ik}, ϖ_1, ϖ_2, ϖ_3, ϖ_{13}, ϖ_{23}, ϖ_{33}, and $\{\varpi_{yd}\}_{d=5}^{9}$, denotes the time-invariant, type-specific, and individual-specific constant. (A large literature focuses on the nonparametric identification of the distribution of random coefficients in the linear regression model; see, for instance, Hoderlein, Klemelä, and Mammen (2010).) Now for any given manager of skill type i, if beliefs are identified by the assignment distribution at entry, by the dynamic pattern of the observed level choices and performance ratings, as argued above, then 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 ϖ_{0i1}, ω_{2i}, and ω_3.

In particular, observe that ϖ_{0i1} and ω_{2i} flexibly capture the non-random variability in wages at each level for individuals with the same priors and observed characteristics and outcomes. Hence, here, as in a standard semiparametric finite mixture model with lognormal components, ϖ_{0i1} and ω_{2i} are identified by the distribution of wages at each level as well as by changes in this level distribution of wages with tenure. Conditional on $\{p_{i1}\}_{i=1}^{4}$ and $\{\alpha_k, \beta_k\}_{k=1}^{3}$, a further source of identification for each ω_{2i} is the time variation in the average wage of managers 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, are continuously employed at the same level but experience different realized performance leading to different posteriors. In principle, the same argument ensures the identification of higher-degree prior terms in the expression for average log wages. In estimation, however, the coefficient ω_3 has proved not significantly different from zero (as well as coefficients on higher-degree prior terms from the learning and human capital premium in wages).

The dummy level parameters $\{\varpi_{0i2}, \varpi_{0i3}\}_{i=1}^{4}$ are identified by the average wage of managers of the same skill type, prior, 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 retained at Level 1. The parameter ω_{111} is identified by the residual variation with tenure of the average wage of managers of the same skill type, prior, age, education, and year of entry if they entered after 1973, continually assigned to Level 1.

The average wage of managers entering the firm in the same year, or before 1974, of the same skill type, prior, and level assignment but 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 wage of individuals of the same skill type, prior, age, education, and level assignment, who entered the firm between 1974 and 1979, compared to those who entered in earlier years, identifies ω_{ym}, $5 \leq m \leq 9$.

Finally, second moments of the distribution of wages at each level for managers with the same age, education, and year of entry in the firm if they entered after 1973, pin down $\{\sigma_{1i}, \sigma_{2i}\}_{i=1}^{4}$ at Levels 1 and 2, and σ_3 at Level 3. Recall that σ_{ik} is the standard deviation of the sum of the productivity shock at the job offered by the second-best firm to managers of type i assigned to job k at my firm, ε_{inkt}, and of measurement error, ε_{inkt}^{m}.

Lastly, note that the implications of my model for wages also yield a set of additional moments for the identification of the distribution of the unobserved state variable, the prior. For instance, the distribution of yearly wage changes among managers continuously assigned to a same level provides additional information about beliefs and $\{\alpha_k, \beta_k\}_{k=1}^{3}$. By treating α_k and β_k as known at Levels 1 and 2, as argued above, from the distribution of observed performance ratings, this wage information provides a direct source of identification for $\Pr(p_{i1})$.

4 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 manager entrants 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, whereas 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, in total 65,851, 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 have 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, producing 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 are assigned to levels, less common titles are 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 appears 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, while 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.

5 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 at each tenure I compute the probability of the observed assignment and wage 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 (unobserved by the econometrician) performance realized in period t for manager n; 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

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

$$= \sum_{i=1}^{I} q_i \left[p_{i1} \Pr(O_{in1}, \ldots, O_{inT_n}|\alpha, i, e_n) + (1 - p_{i1}) \Pr(O_{in1}, \ldots, O_{inT_n}|\beta, i, e_n)\right], \tag{39}$$

where $I = 4$, as discussed in the paper, $\Pr(i|e_n) = q_i$, $\Pr(\alpha|i, e_n, 1) = p_{i1}$, and $\Pr(\beta|i, e_n, 1) = 1 - p_{i1}$. 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 the manager's wage depend on the manager's skill type, also unobserved by the econometrician, computing the likelihood 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 at 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 or beliefs about it. Thus, recorded performance does not provide any additional information about a manager's ability (or skill type) besides the information provided by realized performance. Third, according to the model, because neither the firm nor a manager observe the manager's ability θ, the firm's assignment 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 technological human capital (which is just a function of tenure at the firm and the previous period level assignment), and on the current vector of productivity shocks.

Formally, let $p_{it} = \varphi(p_{i1}|L^o_{in1}, R_{in1}, \ldots, L^o_{int-1}, R_{int-1})$ denote the updated or posterior belief in period t that a manager of skill type i is of high ability, from the prior p_{i1} 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

$$\Pr(L^o_{int}|L^o_{in1}, R^o_{in1}, R_{in1}, \ldots, L^o_{int-1}, R^o_{int-1}, R_{int-1}, \theta, i, e_n)$$
$$= \Pr(L^o_{int}|\varphi(p_{i1}|L^o_{in1}, R_{in1}, \ldots, L^o_{int-1}, R_{int-1}), t-1, L^o_{int-1}), \tag{40}$$

where $\varphi(p_{i1}|L^o_{in1}, R_{in1}, \ldots, L^o_{int-1}, R_{int-1})$ is given by

$$\frac{\alpha^{R_{in1}}_{L^o_{in1}}(1-\alpha_{L^o_{in1}})^{1-R_{in1}} \cdots \alpha^{R_{int-1}}_{L^o_{int-1}}(1-\alpha_{L^o_{int-1}})^{1-R_{int-1}} p_{i1}}{\alpha^{R_{in1}}_{L^o_{in1}}(1-\alpha_{L^o_{in1}})^{1-R_{in1}} \cdots \alpha^{R_{int-1}}_{L^o_{int-1}}(1-\alpha_{L^o_{int-1}})^{1-R_{int-1}} p_{i1} + \beta^{R_{in1}}_{L^o_{in1}}(1-\beta_{L^o_{in1}})^{1-R_{in1}} \cdots (1-p_{i1})}$$

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_{it} 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. Also, recall that the parameters $\{\alpha_k, \beta_k\}^3_{k=1}$ governing the output signals about ability are assumed to be independent of a manager's skill type.

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 at the firm. Thus, I denote the probability density function of the observed wage w^o_{int} in period t for manager n of type i assigned to job L^o_{int} by $f(w^o_{int}|L^o_{int}, p_{it}, i, e_n, t)$.

As for performance ratings, recall that realized performance is unobserved by the econometrician. For the econometrician, the joint likelihood of the observed and true performance ratings of a manager in a period does not depend on the prior about the manager's ability, conditional on the manager's ability. Therefore, for any $t \in \{1, \ldots, T_n\}$, we know that

$$\Pr(R^o_{int}, R_{int}|L^o_{in1}, w^o_{in1}, R^o_{in1}, R_{in1}, \ldots, L^o_{int-1}, w^o_{int-1}, R^o_{int-1}, R_{int-1}, L^o_{int}, w^o_{int}, \theta, i, e_n)$$

$$= \Pr(R^o_{int}, R_{int}|L^o_{int}, t, \theta) = \Pr(R^o_{int}|R_{int}, L^o_{int}, t)\Pr(R_{int}|L^o_{int}, \theta).$$

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

$$\Pr(O_{in1}, \ldots, O_{inT_n}|\theta, i, e_n) = \Pr(L^o_{in1}, w^o_{in1}, R^o_{in1}, \ldots, L^o_{inT_n}, w^o_{inT_n}, R^o_{inT_n}|\theta, i, e_n)$$

$$= \sum_{R_{in1}} \sum_{R_{in2}} \cdots \sum_{R_{inT_n}} \Pr(L^o_{in1}, w^o_{in1}, R^o_{in1}, R_{in1}, \ldots, L^o_{inT_n}, w^o_{inT_n}, R^o_{inT_n}, R_{inT_n}|\theta, i, e_n),$$

which can be also expressed as

$$\Pr(O_{in1}, \ldots, O_{inT_n}|\theta, i, e_n) = \sum_{R_{in1}} \sum_{R_{in2}} \cdots \sum_{R_{inT_n}} \Pr(L^o_{in1}|p_{i1}) f(w^o_{in1}|L^o_{in1}, p_{i1}, i, e_n, 1)$$

$$\cdot \Pr(R^o_{in1}, R_{in1}|L^o_{in1}, 1, \theta) \cdots \Pr(L^o_{inT_n}|\varphi(p_{i1}|L^o_{in1}, R_{in1}, \ldots, L^o_{inT_n-1}, R_{inT_n-1}), T_n - 1, L^o_{inT_n-1})$$

$$\cdot f(w^o_{inT_n}|L^o_{inT_n}, \varphi(p_{i1}|L^o_{in1}, R_{in1}, \ldots, L^o_{inT_n-1}, R_{inT_n-1}), i, e_n, T_n)\Pr(R^o_{inT_n}, R_{inT_n}|L^o_{inT_n}, T_n, \theta). \quad (41)$$

Finally, the sample likelihood is the product of the probabilities in (39) over the N managers:

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

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 maximization problem at each trial parameter vector. The optimization algorithm I use to maximize the likelihood 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 Fortran 90. At the estimated parameter vector, the loglikelihood for the sample is 752.871.[5]

6 Monte Carlo Analysis

The model relies on a multidimensional non-linear 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 based on each simulated dataset. Then, I compare the estimates obtained based on these simulated data with the estimates obtained based 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, obtained from the standard deviation of the estimated parameters over the 50 experiments, as

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

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

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

[5] 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.

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, each obtained from the outer product of the scores of the loglikelihood function. 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 substantively significant are c_{12}, $c_{23} + b_{124}$, $c_{25} + b_{125}$, $c_{26} + b_{125}$, b_{337}, and c_{38}. But for all of these, 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 parameters governing beliefs, the distribution of performance ratings, and job assignment choices might be expected to be difficult. Yet, based on the empirical standard deviations of these parameters across these experiments, it is apparent that most of the model's parameters (the values of which equal the baseline estimates in the second column of Table A.1 plus the biases in the third column) are themselves precisely estimated. Standard errors based on the Hessian matrix are only slightly understated, with the exception of the parameters c_{12}, $c_{23} + b_{124}$, 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.

7 Information Bounds

In the paper I have focused on the implications of my estimates for the one firm for which I have data. Here I argue that I can derive some lower and upper bounds on the informativeness of jobs at my firm's (best) competitor, firm f, as measured by the likelihood ratio of high output between a manager of low and high ability, $\beta_{fk_{ft}}/\alpha_{fk_{ft}}$, based on the estimates of the parameters of the wage process at my firm.

To do this, 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 for a manager's foregone informational and technological human capital, which could have been acquired with employment at the best competitor. In turn, the best competitor's expected output embedded in paid wages is informative about the distribution of (true) performance at the jobs of the best competitor. Recall the notation in the paper; for simplicity, let $y_{fHkt} = y_{fHk}(h_t)$ and $y_{fLkt} = y_{fLk}(h_t)$. The formal result is contained in the following:

Proposition 2. *If $\psi_{i1} \leq 0$ and $y_{fLkt} + \delta\psi_{ik0}/(1-\delta) \geq 0$, then*

$$\frac{\beta_{fk_{ft}}}{\alpha_{fk_{ft}}} \leq \frac{\omega_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1)}{\omega_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1) + \omega_{2i}}. \tag{42}$$

Instead, if $\psi_{i1} \geq 0$ *and* $y_{fLkt} + \delta\psi_{ik0}/(1-\delta) \leq 0$*, then*

$$\frac{\omega_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1)}{\omega_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1) + \omega_{2i}} \leq \frac{\beta_{fk_{ft}}}{\alpha_{fk_{ft}}}. \tag{43}$$

Note that if $\alpha_{fk_{ft}} \geq \beta_{fk_{ft}}$, then the ratio $\beta_{fk_{ft}}/\alpha_{fk_{ft}}$ ranges from zero to one. In practice, based on the parameter estimates, the ratio $\beta_{fk_{ft}}/\alpha_{fk_{ft}}$ appearing on the left-side of (42) and on the right-side of (43) 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,

$$P_{fHk}(p_{it}) = \frac{p_{it}\alpha_{fk_{ft}}}{p_{it}\alpha_{fk_{ft}} + (1-p_{it})\beta_{fk_{ft}}} = \frac{p_{it}}{p_{it} + (1-p_{it})\beta_{fk_{ft}}/\alpha_{fk_{ft}}},$$

so that updated probabilities depend on only the ratio $\beta_{fk_{ft}}/\alpha_{fk_{ft}}$. Thus, based on (42) and (43), 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 firm f rather than at my firm. Starting from an average prior of 0.473 across the four manager skill types (that is, $\sum_i^I q_i p_{i1} = 0.473$ based on the estimates in the paper), I estimate that it would take between 11 and 20 consecutive years of high output at firm f 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 reported 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 2 is as follows. Recall that, by definition, firm f's expected output at state s_{int} and job k_{ft}, per unit of labor input, net of productivity shocks is given by

$$y_f(s_{int}, k_{ft}) = y_{fLkt} + \beta_{fk_{ft}}(y_{fHkt} - y_{fLkt}) + (\alpha_{fk_{ft}} - \beta_{fk_{ft}})(y_{fHkt} - y_{fLkt})p_{it}. \tag{44}$$

Recall that for manager n of type i with observed characteristics $(age_n, edu_n, year_n)$, I specify $y_f(s_{int}, k_{ft})$ as

$$y_f(s_{int}, k_{ft}) = \varpi_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1) + \bar{\omega}_{ik}p_{it}, \tag{45}$$

so from (44) it follows that

$$\varpi_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1) = y_{fLkt} + \beta_{fk_{ft}}(y_{fHkt} - y_{fLkt})$$

and $\bar{\omega}_{ik} = (\alpha_{fk_{ft}} - \beta_{fk_{ft}})(y_{fHkt} - y_{fLkt})$. From the expressions for paid wages in the paper,

$$\ln(w^o_{int}) = \omega_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1) + \omega_{2ik}p_{it} + \omega_3 p_{it}^2 + u_{inkt},$$

$$\omega_{ik}(age_n, edu_n, year_n) = \varpi_{ik}(age_n, edu_n, year_n) + \delta\psi_{ik0}/(1-\delta).$$

As discussed, in estimation ω_{2ik} has proved independent of the job assignment at my firm, k. By definition, $\omega_{2ik} = \bar{\omega}_{ik} + \delta\psi_{ik1}/(1-\delta)$. So both $\bar{\omega}_{ik}$ and ψ_{ik1} can be thought, without loss, as being independent of k. Thus, $\omega_{2ik} = (\alpha_{fk_{ft}} - \beta_{fk_{ft}})(y_{fHkt} - y_{fLkt}) + \delta\psi_{i1}/(1-\delta)$. Simple manipulations yield that

$$\omega_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1) = y_{fLkt} + \beta_{fk_{ft}}(y_{fHkt} - y_{fLkt}) + \delta\psi_{ik0}/(1-\delta), \quad (46)$$

$$\omega_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1) + \omega_{2i} = y_{fLkt} + \alpha_{fk_{ft}}(y_{fHkt} - y_{fLkt}) + \delta(\psi_{ik0} + \psi_{i1})/(1-\delta). \quad (47)$$

From (46) and (47) it also follows that

$$\frac{\beta_{fk_{ft}}}{\alpha_{fk_{ft}}} = \frac{\omega_{ik}(\cdot) + \omega_{1kt}(t-1) - y_{fLkt} - \delta\psi_{ik0}/(1-\delta)}{\omega_{ik}(\cdot) + \omega_{1kt}(t-1) + \omega_{2i} - y_{fLkt} - \delta(\psi_{ik0} + \psi_{i1})/(1-\delta)}. \quad (48)$$

To derive (42), I now use (48) to obtain an upper bound on $\beta_{fk_{ft}}/\alpha_{fk_{ft}}$. Suppose that $\psi_{i1} \leq 0$, so that the compensating wage differential is decreasing in the prior, and suppose that $y_{fLkt} + \delta\psi_{ik0}/(1-\delta) \geq 0$. Thus,

$$\frac{\beta_{fk_{ft}}}{\alpha_{fk_{ft}}} \leq \frac{\omega_{ik}(\cdot) + \omega_{1kt}(t-1) - y_{fLkt} - \delta\psi_{ik0}/(1-\delta)}{\omega_{ik}(\cdot) + \omega_{1kt}(t-1) + \omega_{2i} - y_{fLkt} - \delta\psi_{ik0}/(1-\delta)} \leq \frac{\omega_{ik}(\cdot) + \omega_{1kt}(t-1)}{\omega_{ik}(\cdot) + \omega_{1kt}(t-1) + \omega_{2i}}, \quad (49)$$

where the first inequality follows from $\psi_{i1} \leq 0$ and the second, from $y_{fLkt} + \delta\psi_{ik0}/(1-\delta) \geq 0$.

Now to derive (43), I use (48) to obtain a lower bound on $\beta_{fk_{ft}}/\alpha_{fk_{ft}}$. Suppose that $\psi_{i1} \geq 0$, so that the compensating wage differential is increasing in the prior, and $y_{fLkt} + \delta\psi_{ik0}/(1-\delta) \leq 0$. By inverting (48) I obtain

$$\frac{\alpha_{fk_{ft}}}{\beta_{fk_{ft}}} \leq \frac{\omega_{ik}(\cdot) + \omega_{1kt}(t-1) + \omega_{2i} - y_{fLkt} - \delta\psi_{ik0}/(1-\delta)}{\omega_{ik}(\cdot) + \omega_{1kt}(t-1) - y_{fLkt} - \delta\psi_{ik0}/(1-\delta)} \leq \frac{\omega_{ik}(\cdot) + \omega_{1kt}(t-1) + \omega_{2i}}{\omega_{ik}(\cdot) + \omega_{1kt}(t-1)}, \quad (50)$$

where the first inequality follows from $\psi_{i1} \geq 0$ and the second, from $y_{fLkt} + \delta\psi_{ik0}/(1-\delta) \leq 0$. This completes the proof of the claim.

8 Estimation Including Entrants at Higher Levels

Note that the probability of selection into the sample, equal to the probability of the observed assignment for a manager in the first year of employment at the firm, is determined within the model. Under the model, this probability is a function of prior beliefs, whose distribution is estimated together with the rest of the model's parameters. The assumption implicit in this formulation, and in line with the equilibrium assignment policy implied by 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.* So, in this sense my estimation already takes into account issues of sample selection due to the non-randomness of the data.

However, in order to address potential concerns about selection 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 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 present the estimation results based on the extended sample.

Note that I could have, alternatively, estimated the model on separate samples, corresponding to entrants into the firm at different levels. An argument for such a choice is that all parameters governing job assignment, performance evaluations, and wages may be specific to different types of managers as determined by their entry level. The reason I instead opt for one sample that includes all entrants at Levels 1–4 is to perform a clear-cut 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 flexibility of allowing all parameters to vary across managers depending on their entry level.

8.1 Estimation Sample

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

8.1.1 Sample Construction

As mentioned, the original BGH dataset contains 30,675 observations on entrants into one U.S. firm 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 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 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 further 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 Level 2 and higher, 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 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 individuals 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 3 in the model and ratings of 2, 3, 4, and 5 in the data are reclassified as ratings of zero.

8.1.2 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 6 and 7 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 at the firm at Level 1. The pattern of assignment to the other levels is quite similar, with two main differences. First, the profile of assignment to Level 2 implies that the proportion of managers assigned 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 larger 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 hazard rates, note that, given the absence of demotions, 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 Level 3 are also strikingly similar across the two samples. The hazard rates of retention at Level 3 are very similar, too, across the two samples. The distribution of recorded high ratings at Levels 1 and 2 is also quite similar across the two samples; see Table A.4.

Consider now the distribution of wages in Table A.5. Naturally, the distribution of wages at Level 1 is identical in the two samples. As for the distribution 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 at Level 2 and higher receive on average higher wages compared to entrants at Level 1, even when assigned to the same level in the same tenure year. This is one piece of evidence in support of modeling the existence of persistent differences across managers entering into the firm at different levels. I do so below by allowing for differences in initial priors across managers (of a same skill type) depending on their entry levels.

8.2 Empirical Specification

Here I present the empirical specification of the model, namely, the parameterization of the processes governing initial prior beliefs, level assignments, 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 original sample of entrants at Level 1.

Initial Prior Beliefs Parameters. In specifying the distribution of the initial prior beliefs, I allow for differences in this distribution across the subsample of entrants in the firm at Level 1 and the subsample of entrants at higher levels. I allow for this greater 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 in the paper from the sample of entrants at Level 1. Second, this richer 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 entire real line, to avoid boundary problems in estimation.) However, I now allow the value of each type's initial prior that a manager of that type is of high ability to depend on whether the 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 $\{p_{i1}\}_{i=9}^{12}$. Now, the prior parameters for entrants at Levels 3 and 4 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 assume that $p_{71} = p_{31}$ and $p_{81} = p_{41}$, based on model diagnostics (the Akaike information criterion) and fit. (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.) Thus, estimated prior parameters are p_{11}, p_{21}, p_{31}, p_{41}, p_{51}, p_{61}, q_1, q_2, and q_3.

Productivity and Technological Human Capital Parameters. I assume the same process for productivity and technological human capital acquisition as specified in the paper, that is, the technological human capital acquired by a manager at the firm is just a function of the manager's acquired human capital before entry into the firm, h_1, tenure at the firm, $t-1$, and previous period level assignment, k_{t-1}. Note that my data do not contain direct information about the output of a manager at my firm and beliefs determine both the match surplus value of separation and assignment to the firm's jobs. Hence, as standard, based on the information I have available on managers' job assignments, I can (at most) identify differences between the expected output of a manager at my firm and the expected output of a manager at the second-best firm, the reference alternative. As discussed in the paper, I normalize ν_{0t} and ν_{1t} at zero and, thus, interpret the parameters of expected output as measuring the differences between the magnitude of each such parameter at my firm and the corresponding parameter at the second-best firm.

In light of the flexibility of the productivity and technological 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 in the paper. First, I set to zero any parameters that turn out to be quantitatively insignificant, when constraining them to be equal to zero does not affect any other parameter estimate. Any case in which a productivity parameter equals zero is to be interpreted as a case in which the value of that parameter does not significantly vary across my firm and the second-best firm. Second, when differences in parameters across tenures for the same level or across levels, either for the same tenure or different tenures, are quantitatively insignificant, and have no effect on any other parameter estimate, I set the relevant parameters equal to each other. As in the paper, I allow the slope parameters for one-period expected output to have common components across Levels

1, 2, and 3.

Specifically, for entrants at Level 1, I maintain the same parameterization of expected output as in the paper. 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). Recall that, for the purpose of level assignment and separation, an individual's state variable at the beginning of period t is (p_t, h_t, i), which can be expressed as $(p_t, h_1, t-1, k_{t-1}, i)$ for any $t \geq 2$. Assuming that ℓ captures the effect of h_1 on the productivity and technological human capital process and that, as in the paper, the only relevant dependence on i is through beliefs, I denote expected output at job k at tenure t by

$$y(p_{it}, \ell, t-1, k_{t-1}, k) = a^{\ell}_{kt} + \sum\nolimits_{k'=1}^{3} b^{\ell}_{kk_{t-1}t} I(k_{t-1} = k') + c^{\ell}_{kt} p_{it}.$$

By the discussion above and that in the paper, the parameters a^{ℓ}_{kt}, $b^{\ell}_{kk_{t-1}t}$, and c^{ℓ}_{kt} are to be interpreted as differences between the relevant parameters at my firm and those at the second-best firm, ν_{0t} and ν_{1t}. (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 and the values $v(p_{it}, \ell, t-1, k_{t-1}, k)$, $k = 1, 2, 3$, proved close to linear.) So, any case in which any of the parameters a^{ℓ}_{kt}, $b^{\ell}_{kk_{t-1}t}$, and c^{ℓ}_{kt} are found to be not significantly different from zero, and, hence, set to zero, is to be interpreted as a case in which the corresponding constant terms and slope term for expected output are the same across my firm and the second-best firm. Lastly, since none of the parameters $b^{\ell}_{kk_{t-1}t}$ significantly differ across entrants into the firm at different levels, I set them equal and simply denote them by $b_{kk_{t-1}t}$.

As a result, at Level 1 expected output is given by

$$y(p_{it}, \ell, t-1, k_{t-1}, 1) = a^{1}_{11} I(\ell = 1) I(t = 1) + c^{1}_{12} I(\ell = 1) p_{i2} I(t = 2)$$

$$+[b_{123} I(k_2 = 2) + c^{1}_{13} I(\ell = 1) p_{i3}] I(t = 3) + b_{123} I(k_3 = 2) I(t = 4)$$

$$+b_{125} I(k_4 = 2) I(t = 5) + b_{125} I(k_5 = 2) I(t = 6) + b_{125} I(k_6 = 2) I(t = 7),$$

$1 \leq i \leq 12$, with $a^{1}_{11} = 1{,}000$ (the same normalization as in the paper), b_{123}, $b_{124} = b_{123}$, b_{125}, and $b_{127} = b_{126} = b_{125}$. Thus, estimated parameters at Level 1 are: b_{123}, b_{125}, c^{1}_{12}, and c^{1}_{13}. Next, at Level 2 expected output is given by

$$y(p_{it}, \ell, t-1, k_{t-1}, 2) = a^{2}_{21} I(\ell = 2) I(t = 1) + c^{1}_{22} [I(\ell = 1) + I(\ell = 2)] p_{i2} I(t = 2)$$

$$+c^{1}_{23} [I(\ell = 1) + I(\ell = 2)] p_{i3} I(t = 3) + c^{1}_{24} [I(\ell = 1) + I(\ell = 2)] p_{i4} I(t = 4)$$

$$+[c^{1}_{25} I(\ell = 1) + c^{2}_{25} I(\ell = 2)] p_{i5} I(t = 5) + [c^{1}_{26} I(\ell = 1) + c^{2}_{26} I(\ell = 2)] p_{i6} I(t = 6)$$

$$+[c^1_{26}I(\ell=1)+c^2_{26}I(\ell=2)]p_{i7}I(t=7)+c^1_{28}p_{i8}I(t=8)$$

with $a^2_{21} = 1,000$, $c^2_{2t} = c^1_{2t}$, $2 \le t \le 4$, $c^1_{27} = c^1_{26}$, $c^2_{27} = c^2_{26}$, and $c^3_{28} = c^2_{28} = c^1_{28}$. Omitting level-specific, level-general, and tenure-specific components that have proved to be insignificant, at the estimated parameter values $c^1_{22} = \gamma_{22} - b_{123}$, $c^1_{23} = \gamma_{23} - b_{123}$, $c^1_{24} = \gamma_{24} - b_{123}$, $c^1_{25} = \gamma_{25} - b_{125}$, $c^1_{26} = \gamma_{26} - b_{125}$, and $c^1_{28} = \gamma_{28} - b_{125}$. In practice, γ_{28} has proved negligible. In sum, estimated parameters at Level 2 are γ_{22}, γ_{23}, γ_{24}, γ_{25}, c^2_{25}, γ_{26}, and c^2_{26}. The only common component across Levels 2 and 3 turns out to be γ_{24}. To see this, note that at Level 3 expected output is given by

$$y(p_{it},\ell,t-1,k_{t-1},3)=[c^1_{31}I(\ell=1)p_{i1}+a^3_{31}I(\ell=3)]I(t=1)+c^1_{31}I(\ell=1)p_{i2}I(t=2)$$

$$+[b_{333}I(k_2=3)+c^1_{31}I(\ell=1)p_{i3}]I(t=3)$$

$$+(b_{333}I(k_3=3)+\{c^1_{34}I(\ell=1)+c^2_{34}[I(\ell=2)+I(\ell=3)]\}p_{i4})I(t=4)$$

$$+(b_{335}I(k_4=3)+\{c^1_{35}I(\ell=1)+c^2_{35}[I(\ell=2)+I(\ell=3)]\}p_{i5})I(t=5)$$

$$+(b_{336}I(k_5=3)+\{c^1_{35}I(\ell=1)+c^2_{36}[I(\ell=2)+I(\ell=3)]\}p_{i6})I(t=6)$$

$$+(b_{337}I(k_6=3)+\{c^1_{37}I(\ell=1)+c^2_{37}[I(\ell=2)+I(\ell=3)]\}p_{i7})I(t=7)+c^1_{38}p_{i8}I(t=8)$$

with $a^3_{31} = 1,000$, $b_{334} = b_{333}$, $c^1_{34} = \gamma_{24} + c^1_{31}$, $c^1_{35} = \gamma_{24}$, $c^1_{36} = c^1_{35}$, $c^3_{3t} = c^2_{3t}$ for $4 \le t \le 7$, and $c^3_{38} = c^2_{38} = c^1_{38}$. Differently from the specification estimated in the paper, since c^1_{31} has proved not to be significantly different from zero, I set c^1_{31} equal to zero. Hence, estimated parameters at Level 3 are b_{333}, b_{335}, b_{336}, b_{337}, c^2_{34}, c^2_{35}, c^2_{36}, c^1_{37}, c^2_{37}, and c^1_{38}. Note that expected output at each level in $t = 8$ is specified in the same way as in the paper. (In the first of the two estimated specifications, I also restrict $c^2_{37} = c^1_{37}$ since their difference proved insignificant.)

Exogenous Separation Parameters. To conserve on parameters, I assume that at Level 1 the parameters of the probabilities of exogenous separation satisfy $\eta_{12} = \eta_{11}$ and $\eta_{18} = \eta_{17} = \eta_{16} = \eta_{15} = \eta_{14}$. 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, $\eta_{12} = \eta_{11}$, $\eta_{13} = \eta_{14} + \xi_3$, and $\eta_{18} = \eta_{17} = \eta_{16} = \eta_{15} = \eta_{14}$, and estimated parameters are η_{11}, ξ_3, and η_{14}.) At Level 2, here I assume that $\eta_{22} = \eta_{21}$, $\eta_{24} = \eta_{23}$, $\eta_{26} = \eta_{25}$, and $\eta_{28} = \eta_{27}$. 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, $\eta_{22} = \eta_{21}$, $\eta_{23} = \eta_{22} + \xi_3$, and $\eta_{28} = \eta_{27}$.) At Level 3, I assume that $\eta_{34} = \eta_{33} = \eta_{31}$, $\eta_{37} = \eta_{36} = \eta_{35}$, and $\eta_{38} = 0$. 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, $\eta_{33} = \eta_{32} = \eta_{31}$, $\eta_{3t} = \eta_{2t}$, $4 \le t \le 7$, and $\eta_{38} = 0$.)

Performance Ratings Parameters. I model the process for performance ratings here in the same way as I do in the paper. Thus, here as before, estimated parameters for the true and recorded distribution of ratings are $\{\alpha_k, \beta_k\}^3_{k=1}$ and $(d_0, d_2(1), d_2(2))$.

Wage Parameters. 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, $1 \le i \le 12$ (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

$$\ln(w_{int}^o) = \omega_{ik}(age_n, edu_n, year_n) + \omega_{1kt}(t-1) + \omega_{2i}p_{it} + \omega_3 p_{it}^2 + u_{inkt},$$

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

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

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 $k < 3$.) To avoid parameter proliferation, based on model diagnostics (the Akaike information criterion) and fit, I assume that the coefficient on the first-degree 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 discuss here the two main differences between this specification for wages and that in the paper. The first difference, as just mentioned, is that here I allow the intercept term ϖ_{0i3} at Level 3 and the slope term ω_{2i} to vary across managers entering into the firm at different levels. In terms of ϖ_{0i3}, to conserve on parameters, I impose $\varpi_{053} = \varpi_{013}$ and $\varpi_{093} = \varpi_{013}$ since their differences proved insignificant. In terms of ω_{2i}, to avoid parameter proliferation, I restrict these parameters to be equal across entrants at Levels 2, 3, and 4, as discussed in the previous paragraph. Observe also that since the parameters ω_{2i} prove not to differ substantially across levels, conditional on a manager's skill type/entry level, I also restrict them to be the same across levels as for the specification estimated in the paper. The second difference is that, in light of the additional observations on wages at Level 3 in the new sample, here I let the variance of the random disturbance at Level 3 vary with a manager's skill type. To conserve on parameters, I assume that the variance term σ_{ik} to be 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. In terms of the coefficients on the year-of-entry dummies, I set $\varpi_{ym} = 0$ for $0 \leq m \leq 3$ and $\varpi_{y4} = \varpi_{y5}$, so that estimated parameters are $(\varpi_{y5}, \varpi_{y6}, \varpi_{y7}, \varpi_{y8}, \varpi_{y9})$, as for the specification in the paper. As for the remaining coefficients, here as in the paper, at Level 1 I assume that the coefficient on tenure is $\omega_{11t} = \omega_{111}I(t<5) + \omega_{115}[I(t=5) + (t=6)]$, with $\omega_{115} = -\omega_{111}$. Once more, as in the paper, the coefficients on tenure at Level 2, ω_{12t}, and Level 3, ω_{13t}, prove not significantly different from zero. In estimation, the parameter ω_3 proves not significantly different from zero, as in the paper. As a result, 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_{111}, $\{\omega_{2i}\}_{i=1}^8$, $\{\sigma_{i1}, \sigma_{i2}\}_{i=1}^4$, σ_{13}, σ_{33}, and σ_{43}.

8.3 Estimation Results

I now present and discuss the results of the estimation of my model based 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 assume that $c_{37}^2 = c_{37}^1$, since their difference proves insignificant. (Recall that c_{kt}^{ℓ} denotes the difference in expected output, for managers entering the firm at Level ℓ, between a high and a low ability manager at job k and tenure t.) 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, I simulated 3,000 prior realizations per manager, drawn from the estimated nonparametric distribution of initial priors for each specification.) One difference is that 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 error in recorded performance ratings, are remarkably similar to those in the paper, as discussed below. According to this finding, then, the filtering rules applied to the original data to obtain the estimation sample do not seem to induce appreciable selection.

8.3.1 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 and 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 based on the extended sample successfully captures the tenure profile of separation and assignment to the main jobs 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, in terms of the distribution of managers across levels, in Table A.2, the model tracks the observed distribution remarkably well. In terms of the hazard rates 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 is in 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 each level. 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 similar features 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.[6] Given the larger sample size, almost all parameters are more precisely estimated than when I only use 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 initial priors that a manager is of high ability for the first, second, third, and fourth skill types are given, respectively, by $p_{11} = 0.382$, $p_{21} = 0.372$, $p_{31} = 0.466$, 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 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. Note that for the last two types the estimates of the initial priors 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 original sample prior (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

[6] Recall 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. As before, I follow this convention in reporting the likelihood value.

to 0.381). Entrants at 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 paper's initial priors for the first and second skill types (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 in the paper) with a standard deviation of 0.092 (from $[\sum_i^I q_i(p_{i1} - \sum_i^I q_i p_{i1})^2]^{1/2}$), 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 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 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, analogously to 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 $(1 - \alpha_1)(1 - \beta_3) = 0.248 < (1 - \alpha_3)(1 - \beta_1) = 0.249$, $\alpha_3\beta_2 = 0.267 > \alpha_2\beta_3 = 0.266$, and $(1 - \alpha_3)(1 - \beta_2) = 0.2327 < (1 - \alpha_2)(1 - \beta_3) = 0.2330$.

Finally, the fact that $d_0 = 0.487$, $d_2(1) = -0.668$, and $d_2(2) = -0.525$, compared to the original sample's $d_0 = 0.521$, $d_2(1) = -0.703$, and $d_2(2) = -0.544$, implies comparable estimates for the recording error in performance ratings across the two samples.

8.3.2 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 capture very well 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 this distribution equal to a lower bound on the wages observed 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 discuss model fit and some of the estimates 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 the others. In terms of level assignments, the fit is remarkable. Consider now the hazard rates of separation, retention at level, and promotion at each level. 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 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 the first few years of tenure.

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 initial priors that a manager is of high ability for the first, second, third, and fourth skill types are given, respectively, by $p_{11} = 0.440$, $p_{21} = 0.381$, $p_{31} = 0.465$, 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 $q_4 = 0.321$.

Hence, 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 based on the original sample (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 'innate' abilities of new hires at higher levels vary more than those of entrants at Level 1. My estimates for Specification 2 confirm that 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 $[\sum_i^I q_i(p_{i1} - \sum_i^I q_i p_{i1})^2]^{1/2}$), 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. Finally, the fact that $d_0 = 0.507$, $d_2(1) = -0.693$, and $d_2(2) = -0.539$, whereas the estimates based on the original sample are $d_0 = 0.521$, $d_2(1) = -0.703$, and $d_2(2) = -0.544$, implies similar estimates for the classification error rates in performance ratings across the two samples.

References

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.

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

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.

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.

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

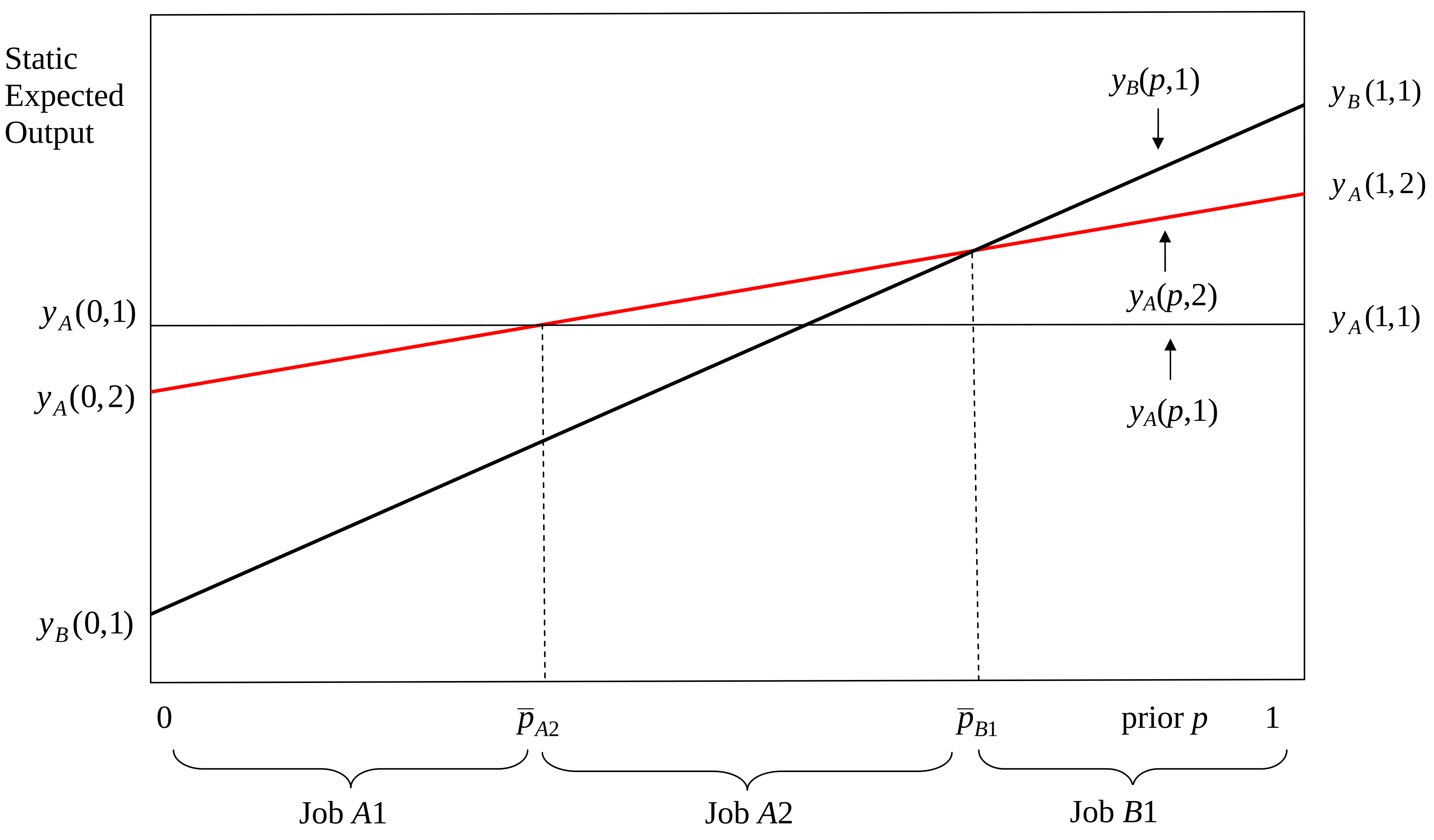


FIGURE 2. Bayesian Updating in Job *A*2

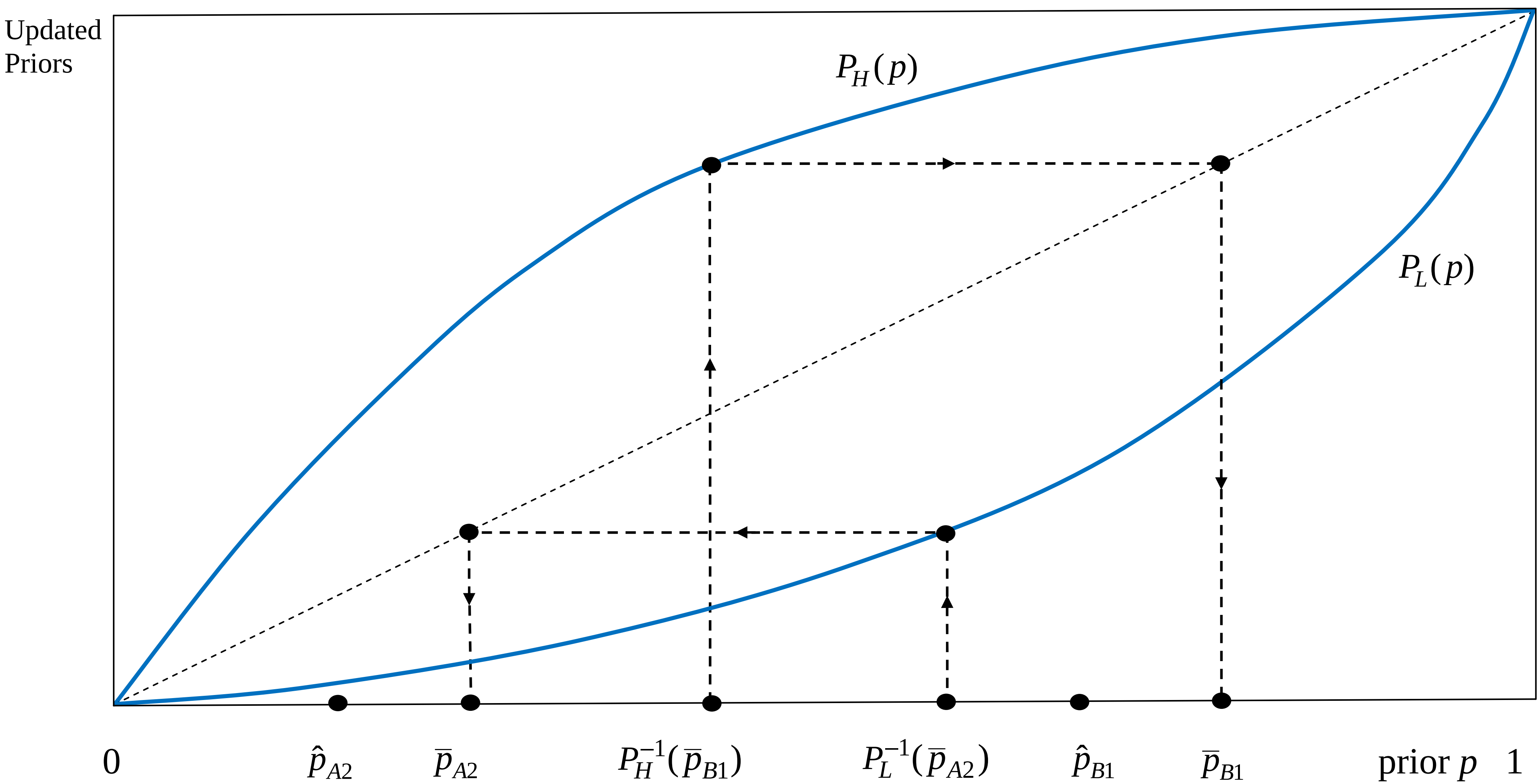
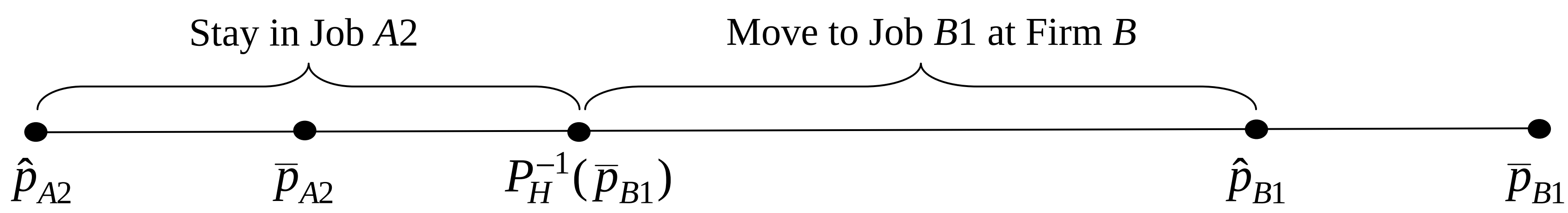
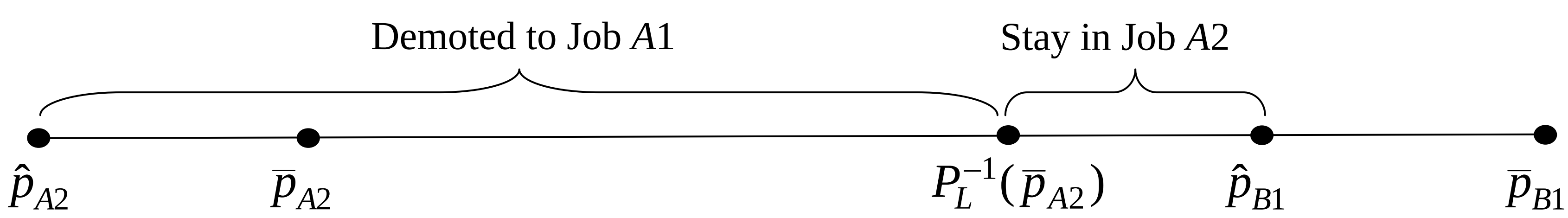


FIGURE 3. Jobs Assigned in Period 2 After Job *A*2 at Firm *A* in Period 1*

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 $\hat{p}_{A2}$ and $\hat{p}_{B1}$ 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
$d_2(1)$	–0.703	0.003	0.395	0.048	0.001
$d_2(2)$	–0.544	–0.001	–0.122	0.033	0.033
Human Capital					
b_{124}	–704.735	–0.837	–0.515	11.493	3.166
b_{125}	–479.607	–0.304	–0.609	3.535	2.823
c_{12}	2,960.515	–3.752	–1.078	24.612	46.467
γ_{22} (= c_{22} + b_{124})	1,858.714	0.381	0.221	12.185	1.930
γ_{23} (= c_{23} + b_{124})	1,505.367	4.036	1.961	14.554	33.448
γ_{25} (= c_{25} + b_{125})	1,629.309	3.025	1.375	15.555	8.282
γ_{26} (= c_{26} + b_{125})	1,745.184	2.476	1.623	10.786	1.959
b_{334}	853.477	1.874	0.588	22.518	16.670
b_{335}	202.791	1.115	1.200	6.570	0.693
b_{337}	228.069	2.124	1.325	11.334	2.620
c_{31}	–399.955	–1.758	–0.885	14.047	10.633
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 ω_{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_{111}	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
$d_2(1)$	–0.668	0.037
$d_2(2)$	–0.525	0.028
Human Capital		
c_{12}^1	2,476.092	16.758
b_{123}	–705.921	5.452
c_{13}^1	2,419.015	8.489
b_{125}	–1,092.881	1.880
γ_{22} (= $c_{22}^1 + b_{123}$)	2,692.672	8.209
γ_{23} (= $c_{23}^1 + b_{123}$)	1,800.283	2.316
γ_{24} (= $c_{24}^1 + b_{123}$)	1,860.990	1.612
γ_{25} (= $c_{25}^1 + b_{125}$)	1,275.078	2.297
c_{25}^2	8,145.406	1.493
γ_{26} (= $c_{26}^1 + b_{125}$)	1,546.718	1.536
c_{26}^2	2,664.069	1.015
b_{333}	1,267.237	1.660
c_{34}^2	1,789.365	2.008
b_{335}	153.565	3.918
c_{35}^2	7,631.714	2.325
b_{336}	180.222	2.053
c_{36}^2	1,879.019	3.449
b_{337}	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 ω_{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_{111}	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
$d_2(1)$	–0.693	0.037
$d_2(2)$	–0.539	0.028
Human Capital		
c_{12}^1	2,546.379	60.027
b_{123}	–921.611	44.749
c_{13}^1	2,654.591	39.086
b_{125}	–1,222.480	42.863
γ_{22} (= $c_{22}^1 + b_{123}$)	2,528.019	47.190
γ_{23} (= $c_{23}^1 + b_{123}$)	1,981.858	118.413
γ_{24} (= $c_{24}^1 + b_{123}$)	1,749.285	5.412
γ_{25} (= $c_{25}^1 + b_{125}$)	1,228.735	7.245
c_{25}^2	4,315.857	29.517
γ_{26} (= $c_{26}^1 + b_{125}$)	1,511.991	20.117
c_{26}^2	5,074.299	71.326
b_{333}	1,586.090	22.701
c_{34}^2	1,713.867	14.304
b_{335}	352.196	50.346
c_{35}^2	3,544.000	645.993
b_{336}	182.064	0.047
c_{36}^2	4,046.057	30.651
b_{337}	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 ω_{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_{111}	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*

Statistic	Wages in Each Case			
		Equal Informativeness as		
	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: $\alpha_k = \hat{\alpha}_3,\ \beta_k = \hat{\beta}_3,\ k = 1, 2$