# Career Dynamics Under Uncertainty: Estimating the Value of Firm Experimentation*

Elena Pastorino
University of Pennsylvania
(JOB MARKET PAPER)
November 2004

**Abstract**

This paper develops and structurally estimates a dynamic model of learning in which a firm can acquire information about a worker's ability by observing his performance over time. Ability determines both the profitability of a job and the job-dependent distribution of performance outcomes. Different output signals about a worker's productivity can be generated by the firm by assigning the worker to different jobs. Because of the trade-off between learning and short-run profit maximization, the firm's optimal information acquisition strategy is the solution to an experimentation problem (a multi-armed Bandit problem with dependent and independent arms). Under the firm's optimal employment policy, the worker is assigned to jobs of decreasing degree of informativeness, as measured by the dispersion in posterior beliefs. The purpose of the analysis is to investigate to what extent uncertainty about ability affects the dynamic pattern of a worker's transition across jobs within a firm, i.e., the timing and job characteristics of a career. To this end the model is structurally estimated using longitudinal data from a single U.S. firm, on the cohorts of managers who enter the firm at the lowest managerial level between 1970 and 1979. Estimation results confirm that a theoretically restricted learning model can succeed in fitting the dynamic profile of the probability of retention and promotion at the major job positions within the firm. The estimated model is then used to compute the firm's value of information and to evaluate the effect on this value, the pattern of job assignments, and on turnover rates of (*i*) changes in the discount rate, which reflect changes in market interest rates, and (*ii*) alternative information structures.

Keywords: Retention, Job Assignment, Learning, Experimentation, Dependent Bandit.
JEL Classification: C73, D21, D83, J41, M12, M51, M54.

---

*I am truly indebted to Kenneth Burdett, Jan Eeckhout, George Mailath, Steven Matthews, Petra Todd and Kenneth Wolpin for their generous advice and encouragement. I benefited from conversations with Antonio Merlo, Nicola Persico, Jesus Fernandez-Villaverde and Iourii Manovskii. I am especially grateful to George Baker for kindly providing the data and to Bengt Holmström for his support in the project. I also wish to thank Hector Chade, Leonardo Felli, David Levine, Dale Mortensen, John Rust, Edward Schlee and the participants in the 2004 North American Summer Meeting of the Econometric Society (Brown University, June 2004), the 2004 Annual Meeting of the Society for Economic Dynamics (Florence, July 2004) and the Fourth Villa Mondragone Workshop in Economic Theory and Econometrics (Rome, July 2004) for their suggestions. Correspondence: Department of Economics, University of Pennsylvania, 160 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104-6297. E-mail: elenap@econ.upenn.edu.

# 1 Introduction

A controversial issue in the theory of the firm is the extent to which promotion and compensation are motivated by a firm's need to provide incentives in the face of moral hazard or to sort employees according to their unobserved abilities. While the existing literature is rich with theoretical contributions, both interpretations have received modest empirical attention (see, for instance, the discussion in Baker and Holmström [1995], Gibbons and Waldman [1999a] and Chiappori [2003]). The understanding of firms' internal organization, and its impact on the allocation of workers to jobs, does nonetheless have important implications for workers' productivity growth with tenure and, therefore, for firms' incentives to employ them.

Intuitively, when a worker's ability is imperfectly observed at the time of hiring, the only way for a firm to assess whether the worker is talented for a job is to employ him and observe his performance over time. However, if the profitability of a job depends on the worker's true skill, then, when deciding whether to employ the worker, or which task to make him perform, the firm has to trade-off the benefit of receiving additional information about his ability against the cost of employing a worker who might be unsuited to the firm's needs. The purpose of this paper is to investigate the role that information acquisition on the part of a firm plays in determining: $(i)$ the ordering of tasks into a hierarchy of job positions, and $(ii)$ the change in a worker's task assignment over his career. Specifically, the focus of the analysis is on quantifying the extent to which uncertainty about ability affects the dynamic pattern of a worker's transition across jobs within a firm, i.e., the timing and job characteristics of a career.

This problem is formalized as a learning game between a firm and a worker. For simplicity, the worker's ability can be one of two levels ('high' or 'low') and it is assumed to be unobserved to both the firm and worker. The firm consists of a finite number of jobs, which differ in their profitability and informational content. In particular, ability is more valuable at jobs which contribute more to the firm's profit. Moreover, since the likelihood of observing any given output realization depends both on the worker's skill and on the job he performs, the revenue realized in a period provides information about the worker's true ability.

The fact that the firm can generate different signals about a worker's productivity by assigning him to different jobs implies that, when allocating a worker to a position, the firm faces the same sequential sampling problem of a decision maker who has to choose one among a given set of alternatives, without knowing the distribution of payoffs associated with each. Because of the implied trade-off between learning and short-run profit maximization, the firm's employment problem can be shown to be strategically equivalent to a particular type of experimentation problem, a Bandit problem with dependent arms (the jobs) and independent arms (the outside option the firm collects if it does not employ the worker). Under suitable restrictions, the solution to this problem is essentially unique and can be completely characterized by a sequence of reservation beliefs. The firm's optimal employment policy prescribes that the worker be assigned to more informative jobs, i.e., those that generate greater dispersion in posterior beliefs, when uncertainty about the worker's human capital is highest, and to more profitable but riskier positions, as the firm learns about the worker's true productivity. In particular, due to the benefit of improved information, it is optimal for the firm to allocate a worker to a job at which he has a strict comparative disadvantage early in his career, when the prior distribution on ability is most diffuse. In this framework, however, learning is typically incomplete, in the sense that the firm always faces (in an ex ante sense) the risk of dismissing a high ability worker, after observing a sequence of low revenue realizations sufficiently long to convince it that the worker's talent is low rather than high.

One purpose of the analysis is to assess the extent to which a learning rationale for job transitions inside a firm can account for the pattern of retention and promotion observed in the data. In order to focus on the interpretation of promotion dynamics as a sorting device, the model intentionally abstracts from issues of incentive provision. As mentioned, promotions could also be rationalized as an incentive mechanism, to induce workers to undertake costly unobserved actions in the interest of the firm. In this case, however, estimation of the effect of screening would require isolating the learning component from the incentive one, given that informational asymmetries arise endogenously in a dynamic moral hazard setup. Therefore, as a first approximation to investigate the empirical relevance of the hypothesis that workers are gradually sorted to higher level jobs, according to their perceived ability, the analysis restricts attention to the problem of information acquisition in a pure learning framework.

To this end the model is structurally estimated, by smooth simulated maximum likelihood, using a unique longitudinal dataset from a single U.S. firm in a service industry between 1969 and 1988. The estimation sample consists of the ten years of observations on job assignments, either Level 1, Level 2 or Level 3, and performance ratings for the cohorts of managers entering the firm at the lowest managerial level, Level 1, between 1970 and 1979, with at least sixteen years of education at entry (i.e., college graduates). The estimation results confirm that the model fits successfully the dynamic profile of the probability of separation from the firm and of retention at Level 1, respectively increasing and decreasing in a worker's tenure. It also captures the qualitative and quantitative features of the pattern of assignment to Level 2, decreasing after the second year since entry at the firm, and to Level 3, at first increasing and then decreasing.

The estimated model can also be used to provide a measure of the firm's value of information and of the inefficiency of job assignment and turnover. Intuitively, since the firm can condition its employment decision in any future period on the performance signal observed in the current period, a natural measure of the gross value of information is the maximal expected extra profit that the firm obtains by observing the worker's output in the current period, due to the improved assessment of his ability. This measure specifically captures the firm's own valuation of the variation in posterior beliefs, i.e., 'new' information, generated by all the possible outcome signals to be realized at each job. In this sense, the firm's demand for information can be uncovered as measured by the firm's willingness to pay to acquire it. Because, as explained, acquiring information about a worker's ability is costly in an opportunity cost sense, the option value of this information can also be quantified and the net value of information estimated. The opportunity cost of information is then measured as the one-period profit loss from choosing the assignment (no employment, Levels 1, 2 or 3) which maximizes dynamic rather than static profit.

Given the estimated values of the parameters of the model, a number of counterfactual exercises are performed. The goal is to investigate the impact on the value of information and, through this, on the probability of retaining a high ability worker (i.e., the extent to which learning takes place through employment), of $(i)$ changes in the firm's degree of time impatience, which reflect changes in market interest rates, and $(ii)$ alternative informational structures. In particular, increased precision of prior information increases the probability of employment of a high ability worker between 1 and 5.4 percent. Compared to the benchmark case, in which parameters are fixed at their estimated values, when Level 1 becomes perfectly informative, i.e., one period of observation of the worker's output at the level perfectly reveals his ability, the value of information to the firm can increase by more than 100 percent. This in turn causes a reduction in the turnover of high ability workers between 30.4 percent, at low tenures, and 3,791.3 percent, at high tenures. The greatest increase in the probability of retention of high ability workers is nevertheless achieved when Level 2 becomes perfectly informative. In this case, the increase in the firm's value of information can be as large as 480.6 percent, down to a minimum of approximately 1 percent at the highest belief values. The corresponding increase in the probability of employment of a high ability worker is between 50.3 and 7,691.9 percent. These results seem to suggest that improved monitoring of workers' performance would be most effective, in terms of the firm's ability to select talented workers, at next-to-entry jobs rather than at entry level positions, given the substantial fraction of exit observed in the data, and predicted by the model, at the intermediate job, Level 2.

The paper is organized as follows. Section 2 introduces the model, Section 3 describes the data and analyzes relevant descriptive statistics. Section 4 presents the solution and the estimation method, while Section 5 contains the estimation results. Section 6 comments on the results of the counterfactual experiments and Section 7 reviews the relevant related literature. Finally, Section 8 briefly concludes and explores directions of further research.

## 2 A Learning Model

Consider a market populated by firms and workers. Time is discrete and has an infinite horizon, with dates $t = 1, 2, ....$ Firms and workers are infinitely-lived and risk-neutral and share the common discount factor $\delta \in [0, 1)$. In what follows the focus is on a particular firm and a potential employee, under the assumption that the revenue generated by the worker at that firm is independent of any other workers' output.

The worker's true ability at the firm is unknown to both the firm and worker. Nevertheless, they both know that this ability can be described by the parameter $\theta$, which can take on only one of two values, high, $\overline{\theta}$, or low, $\underline{\theta}$, where $\overline{\theta} > \underline{\theta}$. The firm and worker' prior distribution at the beginning of period 1 over the worker's unobserved ability is $\Pr(\overline{\theta}) = \phi_1$ and $\Pr(\underline{\theta}) = 1 - \phi_1$, with $\phi_1 \in (0,1)$.

If the firm hires the worker in a period, the worker is assigned to one of three tasks, tasks 1, 2 or 3. Suppose the worker is assigned to task $k$ in period $t$. Focussing on essentials, we assume the revenue generated can be one of two values, $\overline{y}_k$ or $\underline{y}_k$, where $\overline{y}_k > \underline{y}_k$. When the worker's unknown ability is high, revenue is more likely to be high, i.e., $\Pr(\tilde{y}_{kt} = \overline{y}_k | \theta = \overline{\theta}) = \alpha_k$ and $\Pr(\tilde{y}_{kt} = \overline{y}_k | \theta = \underline{\theta}) = \beta_k$, where $1 > \alpha_k > \beta_k > 0$, $k = 1, 2, 3$. In the following we will refer to the task the worker performs in a period equivalently as the job position to which he or she is assigned. We also assume the expected return to the worker outside the match is independent of any knowledge of the worker's ability. Moreover, the worker's ability at the firm is independent of the worker's ability at any other firm.[1]

At the start of any period, the firm proposes employment to the worker at wage $w_t$. If the worker is hired that period, the firm pays the worker the period wage $w_t$ and then allocates him or her to one of the tasks. All the worker does is to either accept the offer made by the firm or reject it.[2] If the firm does not hire the worker, it obtains the period profit $\Pi$ and the worker the period income $U$. At the end of the period, with probability $\xi_k \in (0,1)$ the match dissolves for exogenous reasons, potentially dependent on the task the worker performed. This matching friction can be interpreted either as the probability that the job position to which the worker is assigned is closed, due to adverse market conditions, or as the (reduced-form) probability of a preference shock that forces the worker to leave the firm. In the model, and in estimation, it is meant to capture all instances of separation which do not depend on the worker's ability, as revealed by his performance on the job.

Since the revenue distribution at each job is completely characterized by the worker's unobserved ability, the actual income generated implies that both the firm and worker can update their beliefs about the worker's true talent. Specifically, given the prior $\phi$ at the beginning of period $t$ that the worker's ability is $\overline{\theta}$, and the fact that the worker is assigned to task $k$, the updated posterior, after revenue $\overline{y}_k$ or $\underline{y}_k$ is produced, can be respectively computed as

$$\phi_{kh}(\phi) = \frac{\alpha_k \phi}{\alpha_k \phi + \beta_k (1 - \phi)} \quad \text{and} \quad \phi_{kl}(\phi) = \frac{(1 - \alpha_k)\phi}{(1 - \alpha_k)\phi + (1 - \beta_k)(1 - \phi)}$$

by Bayes' rule. Observe that $\phi_{kh}$ increases in $\phi$ and $\alpha_k$, but decreases in $\beta_k$, while $\phi_{kl}$ increases in $\phi$ and $\beta_k$, but decreases in $\alpha_k$. In particular, $\alpha_k > \beta_k$, $k = 1, 2, 3$, implies $\phi_{kh}(\phi) > \phi_{kl}(\phi)$, for $\phi \in (0,1)$. This implies that observing a high revenue improves on the common assessment of the

[1]Ability, here modelled as firm specific, could also be interpreted as general human capital, as long as the worker's employment history at the firm is unobserved by other firms. For a discussion of the wage dynamics which would emerge in presence of outside labor market competition, see the companion paper. Estimation results contained in the present draft only relate to promotion dynamics, but wage dynamics, in the presence of general human capital, can be accommodated as well. See the discussion in Section 5.

[2]Equivalently, effort can be thought to be verifiable and provided inelastically by the worker, with disutility cost normalized to zero.

worker's true ability. The worker's objective is to maximize his or her expected discounted lifetime income, whereas the firm's is to maximize expected discounted profit.

In the following, without loss of generality, we will restrict attention to Markov Perfect equilibria (MPE's) of the complete information game played by the firm and the worker, for which $\phi$ is the state variable. Actions, histories and strategies can be specified in the usual way. From the assumption that the revenue distribution at each task is Bernoulli, it follows that the updated probability at the beginning of period $t$ that the worker is of high ability, from the sequence of revenue realizations at each task, is a sufficient statistic for the firm's and worker' posterior beliefs. Since all MPE's are essentially time-invariant, the subscript $t$ is omitted and the state will be simply denoted by $\phi$.[3]

Note that if the worker rejects the firm's offer in a period, the firm obtains a flow payoff of $\Pi$, but it does not receive any additional information about the worker's ability. Therefore, if the belief at date $t$ is such that not employing the worker is optimal for the firm, the same choice must be optimal at $t+1$, given that the belief has not changed. Let then $\overline{\Pi} \equiv \Pi/(1-\delta)$ denote the expected discounted profit to the firm if it does not employ the worker. Suppose the firm hires the worker at wage $w_k$ if it employs him or her at task $k$, when $\phi$ denotes the common belief about the worker's ability being high. Let the one period expected revenue at task $k$ be denoted by

$$y_k(\phi) \equiv [\alpha_k \phi + \beta_k(1-\phi)]\overline{y}_k + [(1-\alpha_k)\phi + (1-\beta_k)(1-\phi)]\underline{y}_k.$$

In this case the expected return to the firm, from assigning the worker to task $k = 1, 2, 3$, can be expressed as

$$\widetilde{\Pi}_k(w_k, \phi) = y_k(\phi) - w_k(\phi) + \delta(1-\xi_k)E_k[\Pi(\tilde{\phi})|\phi] + \delta\xi_k\overline{\Pi}$$

where the expectation $E_k$ is taken over the future values of the posterior, $\tilde{\phi}$, conditional on its current period value $\phi$ and the task $k$ the worker performs in the period, and $\Pi(\cdot)$ denotes the firm's maximal value from the problem. Notice that the firm and the worker will meet in the following period with probability $1-\xi_k$.

The wage paid by the firm is relatively simple to derive. Let $\overline{U} \equiv U/(1-\delta)$ denote the worker's expected discounted lifetime income. The worker will accept employment at the firm in a period if and only if $V_k(\phi) \geq \overline{U}$, where $V_k(\phi)$ denotes the worker's expected discounted lifetime income if assigned to task $k$, when the firm and the worker's belief is $\phi$. It can be shown that $V_k(\cdot)$ is strictly increasing in the wage paid, for any $k = 1, 2, 3$ and $\phi$. As a consequence, the firm will maximize its expected return if it hires the worker at wage $z$ such that $V_k(z) = \overline{U}$. In particular, in equilibrium the worker is paid $U$ in any period of employment.[4] Hence, for $k = 1, 2, 3$, $\Pi_k(\phi) = \max_w \widetilde{\Pi}_k(w_k, \phi) = \Pi(U, \phi)$,

[3]For the characterization of equilibria, see the companion paper. Note that the perfection requirement reduces the equilibrium set to stationary equilibria which are essentially unique in the outcome of interest, i.e., sample paths along which the worker is continuously employed at the firm.

[4]This is an immediate consequence of the fact that the worker's best response consists in accepting any wage offer at least equal to $U$ and rejecting any other offer.

where

$$
\begin{aligned}
\Pi_k(\phi) &= y_k(\phi) - U + \delta(1-\xi_k)E_k[\Pi(\tilde{\phi})|\phi] + \delta\xi_k\overline{\Pi} \\
&= p_k(\phi)[\overline{y}_k - U + \delta(1-\xi_k)\Pi(\phi_{kh}(\phi))] \\
&\quad +(1-p_k(\phi))[\underline{y}_k - U + \delta(1-\xi_k)\Pi(\phi_{kl}(\phi))] + \delta\xi_k\overline{\Pi}
\end{aligned}
$$

and $p_k(\phi) \equiv \alpha_k\phi + \beta_k(1-\phi)$ is the probability that high revenue realizes when the worker performs task $k$.

If the firm hires the worker in a period, given belief $\phi$, it maximizes its expected return by assigning the worker to task $j$, where $\Pi_j(\phi) \geq \Pi_k(\phi)$, $j, k = 1, 2, 3$. Further, the firm employs the worker if and only if $\Pi_j(\phi) \geq \overline{\Pi}$. In particular, the firm's value function $\Pi(\cdot)$ satisfies the following Bellman equation,

$$
\begin{aligned}
\Pi(\phi) &= \max\{\overline{\Pi}, y_1(\phi) - U + \delta(1-\xi_1)E_1[\Pi(\tilde{\phi})|\phi] + \delta\xi_1\overline{\Pi}, \\
&\quad y_2(\phi) - U + \delta(1-\xi_2)E_2[\Pi(\tilde{\phi})|\phi] + \delta\xi_2\overline{\Pi}, \\
&\quad y_3(\phi) - U + \delta(1-\xi_3)E_3[\Pi(\tilde{\phi})|\phi] + \delta\xi_3\overline{\Pi}\}.
\end{aligned}
$$

The difference between the firm's expected discounted profit from assigning the worker to task $k$ and to task $k'$, $k, k' = 1, 2, 3$ and $k' \neq k$, can be expressed as

$$
\begin{aligned}
\Pi_k(\phi) - \Pi_{k'}(\phi) &= y_k(\phi) - y_{k'}(\phi) \\
&\quad +\delta\{(1-\xi_k)E_k[\Pi(\tilde{\phi})|\phi] - (1-\xi_{k'})E_{k'}[\Pi(\tilde{\phi})|\phi] + (\xi_k - \xi_{k'})\overline{\Pi}\}. \qquad (1)
\end{aligned}
$$

The sign of the difference $\Pi_k(\phi) - \Pi_{k'}(\phi)$ therefore depends on the magnitude of the difference in the one period expected revenue, the first term in (1), and in the expected continuation profit, the second term in (1), between the two tasks $k$ and $k'$. In fact, at any state the return to the firm from task $k$ can be decomposed in the expected revenue produced by the worker in the period and in the expected continuation value, which depends on the additional information about the worker's ability conveyed by the revenue realized. Since the firm's value function is convex in the posterior belief, as is proved below, this information is of value as long as there is uncertainty about the worker's true ability, i.e., $E_k[\Pi(\tilde{\phi})|\phi] \geq \Pi(\phi)$.

In general, the difference $\Pi_k(\phi) - \Pi_{k'}(\phi)$ depends on the particular configuration of parameter values. For instance, suppose $\alpha_1 = 0.900$, $\alpha_2 = 0.828$, $\alpha_3 = 0.999$, $\beta_1 = 0.069$, $\beta_2 = 0.000$, $\beta_3 = 0.274$, $\overline{y}_1 = 114.982$, $\overline{y}_2 = 1,601.966$, $\overline{y}_3 = 4,453.495$, $\underline{y}_1 = -1,998.849$, $\underline{y}_2 = -5,119.335$, $\underline{y}_3 = -369,085.007$, $\xi_1 = \xi_2 = \xi_3 = 0.000$ and $\delta = 0.95$. At all belief values between $\phi = 0.033$ and $\phi = 0.420$, where task 1 is more profitable than task 2 (i.e., $\Pi_1(\phi) - \Pi_2(\phi) \geq 0$), $y_1(\phi) > y_2(\phi)$ holds true but $E_1[\Pi(\tilde{\phi})|\phi] - E_2[\Pi(\tilde{\phi})|\phi] < 0$. On the other hand, if $\delta = 0.99999$, between $\phi = 0.341$ and $\phi = 0.420$ there exist values of $\phi$ for which task 2 is more profitable than task 1 (i.e., $\Pi_2(\phi) - \Pi_1(\phi) \geq 0$), so that the positive difference $E_2[\Pi(\tilde{\phi})|\phi] - E_1[\Pi(\tilde{\phi})|\phi]$ offsets the negative difference $y_2(\phi) - y_1(\phi)$. Moreover, when $\alpha_k = \alpha_{k'}$, $\beta_k = \beta_{k'}$ and $\xi_k = \xi_{k'}$, $k, k' = 1, 2, 3$ and $k' \neq k$, it follows that $E_k\Pi(\phi) = E_{k'}\Pi(\phi)$, since the distribution of the updated posterior is the same at tasks $k$ and $k'$.[5] Hence, to make further progress, additional restrictions have to be imposed.

The main assumption we formulate on the profitability of the three tasks is the following:

$$\begin{aligned} &(A1): \quad y_3(\overline{\theta}) > y_2(\overline{\theta}) > y_1(\overline{\theta}), \;\; y_1(\underline{\theta}) > y_2(\underline{\theta}) > y_3(\underline{\theta}) \\ &(A2): \quad y_3(\overline{\theta}) > \Pi + U > y_1(\underline{\theta}) \end{aligned}$$

where $y_k(\theta) \equiv E(y_k|\theta)$ is the one period expected revenue to the firm from assigning the worker to task $k$ in period $t$, conditional on his or her ability being $\theta$.[6] Assumption (A1) is meant to capture the feature that the impact of ability on expected revenue is greater at potentially more profitable tasks. This restriction also implies that task $y_2$ entails the risk of greater output destruction than task 1, if the worker assigned to it is not of high ability. Similarly, task 3 is 'riskier' than task 2 in output terms.[7] As for (A2), the assumption $y_3(\overline{\theta}) > \Pi + U$ ensures that employment can be profitable for the firm, while the restriction $\Pi + U > y_1(\underline{\theta})$ implies that the firm might find optimal not to hire the worker than to employ him or her at any task. In particular, the firm would never hire a worker of low ability, if it could perfectly observe $\theta$. The first result can then be proved.

**Proposition 1.** *The firm's value function* $\Pi(\cdot)$ *is well-defined, continuous and convex. Under (A1) and (A2), it is also increasing.*

**Proof**: See Appendix A. □

Intuitively, characterizing the firm's optimal retention and task assignment policy requires comparing the maximal expected profit that the firm could obtain from assigning the worker to each of the three tasks. As discussed, the sign of the difference $\Pi_k(\phi) - \Pi_{k'}(\phi)$, $k, k' = 1, 2, 3$ and $k' \neq k$, depends in turn on the difference in the expected one period revenue and in the expected continuation profit from tasks $k$ and $k'$. In particular, even if the firm's value function was strictly convex, the difference between the expected discounted profit from tasks $k$ and $k'$ could be non monotonic.[8] By assumptions (A1) and (A2), however, the difference in the one period revenue from any two tasks is strictly monotonic in $\phi$. Namely, the difference $y_k(\phi) - y_{k'}(\phi)$, $k > k'$, is strictly increasing. Then, in the static case, the unit interval can be partitioned in regions where task $k$ is unambiguously preferred to task $k'$ or viceversa. This observation suggests that a set of sufficient conditions for a characterization of the firm's employment policy can be identified by guaranteing that a global monotonicity condition holds for the difference $\Pi_k(\phi) - \Pi_{k'}(\phi)$.

[5]See Subsection 4.1 for a description of the numerical solution method.

[6]These restrictions will not be imposed in the estimation of the model. See the discussion in Section 5.

[7]Observe that $y_k(\overline{\theta}) > y_k(\underline{\theta})$, $k = 1, 2, 3$, is, instead, a consequence of the fact that $\alpha_k > \beta_k$ implies that the revenue distribution at task $k$, when the worker is of high ability, first-order stochastically dominates the revenue distribution at the same task, when he is of low ability.

[8]Strict convexity can be shown to hold if the expected one period revenue at each task is strictly convex in $\phi$. One way would be to assume that the firm incurs a one period stochastic cost of supervision, as a fraction of the revenue produced, in monitoring the worker's performance at any task and that this cost depends on the worker's true ability.

Under some conditions, it can be shown that the single-crossing property of the static revenues $y_k(\phi)$ and $y_{k'}(\phi)$, $k \neq k'$, implied by (A1) and (A2), translates into an analogous single-crossing property of the dynamic profits $\Pi_k(\phi)$ and $\Pi_{k'}(\phi)$. Specifically, let $\phi_{0,1}$ be the cut-off belief value which makes the firm indifferent between not hiring the worker and employing him at task 1 in the static case, when $\delta = 0$. Similarly, let $\phi_{k,k+1}$, $k = 1, 2$, be the cut-off belief which makes the firm indifferent between tasks $k$ and $k+1$ when $\delta = 0$. Condition $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$ guarantees that the belief values for which the firm is indifferent, respectively, between not employing the worker and allocating him to task 1 ($\phi_{0,1}$), between tasks 1 and 2 ($\phi_{1,2}$) or between tasks 2 and 3 ($\phi_{2,3}$) can be ordered. Then, the firm's policy in the static problem consists in assigning the worker to task 1 if $\phi \in [\phi_{0,1}, \phi_{1,2})$, to task 2 if $\phi \in [\phi_{1,2}, \phi_{2,3})$, to task 3 if $\phi \in [\phi_{2,3}, 1]$ and not employing him or her altogether otherwise.[9] As for the comparison of the expected continuation values, whenever the distribution of the updated posterior at task $k$ is a mean-preserving spread of the corresponding distribution at task $k+1$, i.e., task $k$ is more informative about ability than task $k+1$, it follows $E_k\Pi(\cdot) \geq E_{k+1}\Pi(\cdot)$, with $\Pi(\cdot)$ increasing and convex. This is a consequence of the fact that, being the firm's uncertain about the worker's true worth, it values dispersion is posterior beliefs. Then, more informative tasks, which cause a greater spread in the distribution of the updated posterior, are those which are more profitable when the prior distribution is most diffuse.

The restriction $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$ also implies that there might exist a range of belief values for which the worker is assigned to task 2 in equilibrium, and task 2 is preferred to task 3, even in the dynamic case, if $(i)$ there exists an interval of beliefs for which task 2 is statically more profitable than task 3, past the static cut-off $\phi_{1,2}$, and $(ii)$ task 2 is more informative than task 3. The reason is that, due to the greater informativeness of task 2 as compared to task 3, the threshold belief which makes the firm indifferent between tasks 2 and 3 in the dynamic case, $\phi_2^*$, is typically greater than $\phi_{2,3}$. However, $y_3(\overline{\theta}) > \Pi + U$ implies that when $\phi$ is sufficiently close to 1, task 3 is the dominant choice for the firm. Then, only if $\phi_2^*$ is smaller than $\phi_3^*$, the cut-off belief for which the firm is indifferent between tasks 2 and 3 in the dynamic case, the firm benefits from assigning the worker to task 2, when $\delta > 0$.

Define $\overline{\phi}$ to be the belief which makes the firm indifferent between tasks 1 and 2, whenever task 1 is perfectly informative about ability, while task 2 does not provide any information about the worker's true skill. Given the trade-off between the additional payoff generated at task 2, if the worker's assessed ability is sufficiently high, and the greater informativeness of task 1, $\overline{\phi}$ is indeed an upper bound on the range of beliefs for which the firm might prefer assigning the worker to task 1 rather than to task 2 in the static case. It follows

$$\overline{\phi} \equiv \frac{y_1(\underline{\theta}) - y_2(\underline{\theta}) + \frac{\delta(1-\xi_2)\Pi}{1-\delta(1-\xi_2)}}{\frac{y_2(\overline{\theta})-y_2(\underline{\theta})}{1-\delta(1-\xi_2)} - y_1(\overline{\theta}) + y_1(\underline{\theta}) - \frac{\delta(1-\xi_1)(y_3(\overline{\theta})-U-\Pi)}{1-\delta(1-\xi_3)}}.$$

[9] Observe that we assumed that, whenever indifferent, the firm assigns the worker to the task at which the impact of ability on expected revenue is highest. No employment is meant to indicate all those instances in which the firm offers a wage strictly smaller than $U$.

Note that $\overline{\phi} \in (0,1)$ as long as $\xi_k$, $k = 1,2,3$, is sufficiently small. Let also $k(\overline{\phi}) \equiv \overline{\phi}/(1-\overline{\phi})$. The formal characterization result of the firm's employment policy is contained in the following Proposition.

**Proposition 2.** *Let (A1) and (A2) hold. Suppose* $y_1(\overline{\theta}) > \Pi$, $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$, $\alpha_1 \geq \alpha_2 \geq \alpha_3$ *and* $\beta_3 \geq \beta_2 \geq \beta_1$. *Then, there exists* $\{\underline{\xi}_k, \overline{\xi}_k\}_{k=1}^3$, *with* $0 < \underline{\xi}_k < \overline{\xi}_k < 1$, *such that* $\xi_k \in (\underline{\xi}_k, \overline{\xi}_k)$, $k = 1,2,3$, $\xi_3 \geq \xi_2 \geq \xi_1$ *and* $y_2(\underline{\theta}) - y_3(\underline{\theta}) > k(\overline{\phi})[y_3(\overline{\theta}) - y_2(\overline{\theta})]$. *In this case,* $0 < \phi_1^* < \phi_2^* < \phi_3^* < 1$ *exist such that in any MPE the firm's essentially unique employment policy consists in not employing the worker if* $\phi \in [0, \phi_1^*)$, *assigning him or her to task 1 if* $\phi \in [\phi_1^*, \phi_2^*)$, *to task* 2 *if* $\phi \in [\phi_2^*, \phi_3^*)$ *and to task* 3 *if* $\phi \in [\phi_3^*, 1]$. *Moreover,* $\phi_1^* < \phi_{0,1}$, $\phi_2^* > \phi_{1,2}$ *and* $\phi_3^* > \phi_{2,3}$.

**Proof**: See Appendix A. □

The set of conditions listed in the Proposition guarantee that the same qualitative features of the optimal policy in the static case carry over to the dynamic case.[10] In particular, the firm's optimal employment policy is again an interval belief strategy, with increasing cut-offs determined by the points of indifference between the alternative-specific values $\Pi_k$'s, i.e., the expected discounted profit to the firm from assigning the worker to task $k$. Notice that, modulo the way indifference is solved, the firm's employment policy is also uniquely determined, given that, from single-crossing, the differences $\Pi_1(\phi) - \Pi_2(\phi)$ and $\Pi_2(\phi) - \Pi_3(\phi)$ are strictly decreasing in $\phi$, so that the cut-offs $\phi_k^*$, $k = 1,2,3$, are unique. Also, the result that $\phi_1^* < \phi_{0,1}$ and $\phi_2^* > \phi_{1,2}$ implies that the worker is assigned to task 1 for belief values for which, in the static case, respectively, either employment would not be profitable or task 2 would be more profitable than task 1. Similarly, from $\phi_3^* > \phi_{2,3}$ it follows that task 2 is allocated to the worker over a belief range for which, in the static case, the firm would make the worker perform task 3 rather than task 2. This distortion in the dynamic cut-offs, with respect to the static threshold beliefs, implies that it is optimal for the firm to distort the pattern of static comparative advantage, to generate information about the worker's ability when the worker's true worth is uncertain.

Finally, the conditions under which the characterization result in Proposition 2 holds have also implications for the probability of retention of a high ability worker. As expected, the possibility for the firm to experiment on the worker's ability at tasks which are more informative than task 3 reduces the probability of inefficient turnover of high ability workers. The following result can then be proved.

**Proposition 3.** *Let (A1) and (A2) hold. Suppose* $y_1(\overline{\theta}) > \Pi$, $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$, $\alpha_1 \geq \alpha_2 \geq \alpha_3$ *and* $\beta_3 \geq \beta_2 \geq \beta_1$. *Then, there exists* $\{\underline{\xi}_k, \overline{\xi}_k\}_{k=1}^3$, *with* $0 < \underline{\xi}_k < \overline{\xi}_k < 1$, *such that* $\xi_k \in (\underline{\xi}_k, \overline{\xi}_k)$, $k = 1,2,3$, $\xi_3 \geq \xi_2 \geq \xi_1$ *and* $y_2(\underline{\theta}) - y_3(\underline{\theta}) > k(\overline{\phi})[y_3(\overline{\theta}) - y_2(\overline{\theta})]$. *In equilibrium in the long run only a high ability worker is retained by the firm and employed at task 3. Moreover, the probability of*

[10] Notice that the restrictions on $\xi_k$, $k = 1,2,3$, would reduce to $\xi_3 \geq \xi_2 \geq \xi_1$ if the firm's outside option, $\Pi$, was zero.

*permanent retention of a high ability worker, strictly smaller than one, is higher when at least tasks* 1 *or* 2 *and task* 3 *are assigned in equilibrium than when only task* 3 *is.*

**Proof**: See Appendix A. □

The proposition implies that the assignment of tasks 1 and 2 in equilibrium has merely a screening purpose, and it is optimal as long as there is uncertainty about the worker's ability. As characteristic of experimentation problems, also, limiting learning is incomplete. The firm, in an ex ante sense, always faces the risk of observing a sequence of low output realizations sufficiently long to convince it that the worker is actually of low ability, even if his or her true ability is high. One of the purposes of the empirical analysis is indeed to assess the extent to which changes in the informational structure can improve on the firm's capacity to identify high ability workers, by observing their performance at different jobs. This in turn requires investigating the effect on the profitability of employment of changes in the firm's valuation of information on ability. Measurement and estimation of the value of experimentation are discussed in more detail in Section 6.

# 3 Data

## 3.1 Sample and Variable Definitions

The data consist of personnel records for all management employees of a medium-sized U.S. firm in a service industry between 1969 and 1988. As described in Baker, Gibbs and Holmström [1994a] (BGH), these records include information on every managerial employee in the firm as of December 31 of each year. Each record consists of an employee ID number, the employee's year of entry, age, education, job title and level, cost center code (i.e., the six-digit code of the organizational unit defined for measuring costs, revenues or profits), salary, salary grade (available from 1979 to 1988), bonus and a job performance rating (from 1, lowest, to 5, highest). In total the data contain 74,071 observations on managerial employees at the firm over the sample years. Salary, title and performance rating are year-end values. It is unclear though when, during the year, pay or title changes occurred or performance ratings were attributed, so these variables may not be exactly concurrent. In the empirical analysis we assume, consistently with the model, that title changes occurred after performance ratings were recorded. However, titles were not coded for some new hires in the last years. Specifically, missing data are significant in 1987 and 1988, in which approximately 10 percent of employees and half of new hires do not have title data.

The size of entry cohorts into managerial positions at the firm grew significantly during the sample period. The entry cohort in 1970 was 230 individuals, while by 1988 it was 1175. BGH report that management constituted about 20 percent of total employment each year. The average age of employees entering managerial positions was 33 with a standard deviation of 8 years; the range was from a minimum of 20 to a maximum of 71 years. The average number of years of education was 15.6 with a standard deviation of 2.4 years; the range was from a minimum of 12 to a maximum of

23 years. Both age and education show little variation across cohorts.[11] As for exit, for the sample of entrants at the firm between 1970 and 1979, 10.9 percent left the firm after one year, while 20.4 percent left after two years and 57.7 percent after nine years.[12]

BGH aggregated job titles into levels according to the timing and frequency of transitions of employees across titles. Specifically, as explained in BGH, in the original data there were 276 different titles, but 14 titles, each representing at least 0.5 percent of employee-years, comprised about 90 percent of the observations and 93 percent of those in which the title was coded. In order to fill the job ladder to the top of the firm's hierarchy, BGH added the top title of Chairman-CEO, together with the only two titles observed in moves from the fourteen major titles to the position of Chairman. Transition matrices were then constructed to analyze movements of employees between these seventeen titles, both for individual years and over the whole sample.

Eight job levels were constructed. Level 1 consists of the three titles which employed almost only new hires. Most moves from Level 1 within the firm were to six other titles, identified as Level 2. Moves from Level 2 were almost exclusively to three other job titles, categorized as Level 3. This process was continued until the original seventeen titles were assigned to 8 job levels, with Chairman-CEO at Level 8. After major titles were assigned, less common titles were assigned to levels based on moves between them and titles already assigned.

The hierarchy which emerges from this level structure consists of two parts, Levels 1-4 and Levels 5-8, with Levels 1-4 containing 97.6 percent of employees, each of approximately the same size. Specifically, over the sample period 16,981 employees are at Level 1, 17,725 at Level 2, 17,253 at Level 3, 13,892 at Level 4. The corresponding figures at Levels 5-8 are 1,194, 373, 56 and 20. It is commonly interpreted that upper level jobs correspond more to general management, while lower level jobs depend more on specialized functional knowledge and require performing less complex tasks. For instance, as described by BGH, at Levels 1-4 about 60 percent of the jobs correspond to specific 'line' (revenue-generating) business units, positions with direct contact with customers or creating and selling products, while 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, respectively, to 45 and 25 percent, while general management descriptions such as 'General Administration' or 'Planning' increase to about 30 percent. At Levels 7-8 all jobs are of this form and they entail managing large groups, coordinating across business units and strategic planning, responsibilities which possibly rely more on firm-specific rather than general skills. These observations suggest that the task content of higher level jobs is consistent with our assumption that human capital is most valuable at those jobs.

Over the twenty year sample, the firm has been remarkably stable in the composition of titles and

[11] The composition of entrants across job titles did not change markedly, though there was a relative increase in lower level entry during the years 1976-1985. BGH report that the proportion of minorities and women increased steadily. Our data, though, do not include information on sex or race.

[12] As noted by BGH, patterns are similar for later entrants, even if the average career length becomes shorter over time.

levels. Even as firm size has tripled, the fraction of people at each level has changed very little. After 1984, some new titles were created, but only two are of significant size, representing respectively only 0.6 percent and 0.9 percent of employees (see Table 1 in BGH).

In the data, it is not possible to distinguish whether new entrants into managerial positions in any given year are also new hires at the firm. For instance, a worker could have been promoted from a clerical to a management position. Because a promotion in this case entails a major shift in job tasks, as argued in BGH, and a change from hourly to salaried employment, new promotees into managerial positions are likely to be treated similarly to outside hires. In estimation we focus on the individuals who entered managerial positions between 1970 and 1979 at Level 1. Each entrant cohort is followed for 10 years. This restriction reduces the original sample of 16,133 individuals to 2,714 individuals. The estimation sample is further restricted to the 1,552 individuals with 16 or more years of education at entry.

Performance ratings were coded in the data from 1 (best) to 5 (worst). Ratings of 3, 4 and 5 comprise only a small fraction of all ratings. Ratings 2 to 5 where therefore combined into a single rating, leading to a binary classification of 1 (high rating) and 0 (low rating) as in the model.[13]

### 3.2 Descriptive Statistics

The model has implications for the ex ante probability that in each period a worker assigned at entry to Level 1 will remain at Level 1 or will be assigned to Level 2 or 3, or will leave the firm. Table 1 shows the proportion of employees at each level, as well as the proportion who separated for each year since entry over a ten year period, for the sample of employees entering the firm between 1970 and 1979 at Level 1 with at least 16 years of education and no level information missing.[14]

As noted, at entry all employees are at Level 1. In the second period, 39.6 percent of employees who entered the firm are assigned to Level 2 and 13.3 percent leave. In the third period, only 17.6 percent of the individuals assigned to Level 1 in the first period remain at Level 1, while 47.6 percent are at Level 2 and 9.3 percent at Level 3. The fraction of employees in Level 1 and 2 jobs rapidly decreases with tenure at the firm, while the proportion of employees assigned to Level 3 increases

[13]Ratings of 1 and 2 represent 80.5 percent of all the ratings observed in the original sample (28,398 employee-years have missing rating information, where only 4,703 individuals having no missing rating information in any period) and 89.5 percent of the ratings in the sample used in estimation, in which, by construction, no rating information is missing (see Table B1 in Appendix B). To preserve the informativeness of observed performance about employees' productivity in a year, a rating of 1 has been treated a success, while a rating of $2, 3, 4$ and 5 as a failure. See Appendix B for a comparison of the fraction of ratings 1 through 5 is the original sample and in the estimation sample.

[14]The corresponding statistics for the current estimation sample of 502 individuals, with at least 16 years of education at entry and no level or performance rating missing, are reported in Appendix B. However, for this sub-sample only 22 employees are observed at Level 4 (24 in the sample of 698 individuals which include all education groups) and none at Levels 5 through 8. Observations at Level 4 were therefore added to Level 3. For the total of 1,552 managerial employees with at least 16 years of education at entry, over the first ten years there are only 1,359 observations on employees at Level 4 (11.4 percent of all observations), 15 on employees at Level 5 and 8 on employees at Level 6. Observations on Level 4 to 6 have similarly been added to observations on Level 3.

until the fifth year after entry and then decreases. The proportion of workers who have left the firm is substantial in each year. By the last period of observation, 66.8 percent of the individuals hired at Level 1 have left the firm.

**Table 1. Distribution of Employees Across Levels (16 or More Years of Education at Entry, Missing Ratings - 1,552 Employees)**

| Years Since Entry | Level 1 | Level 2 | Level 3 | Exit | Total |
|---|---|---|---|---|---|
| 0 | 1.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| 1 | 0.471 | 0.396 | 0.000 | 0.133 | 1.000 |
| 2 | 0.176 | 0.476 | 0.093 | 0.256 | 1.000 |
| 3 | 0.079 | 0.309 | 0.261 | 0.352 | 1.000 |
| 4 | 0.047 | 0.193 | 0.327 | 0.434 | 1.000 |
| 5 | 0.028 | 0.128 | 0.352 | 0.492 | 1.000 |
| 6 | 0.020 | 0.085 | 0.351 | 0.544 | 1.000 |
| 7 | 0.016 | 0.066 | 0.329 | 0.588 | 1.000 |
| 8 | 0.011 | 0.050 | 0.311 | 0.628 | 1.000 |
| 9 | 0.008 | 0.037 | 0.286 | 0.668 | 1.000 |

The hazard rates of employment termination and promotion are displayed in Table 2, stratified by tenure at each level. At Level 1 the separation hazard is approximately constant over time at about 0.1. The hazard rate for promotions to Level 2 increases in the second year of tenure in Level 1 (from 0.396 to 0.486), it follows slightly in year 3 to 0.436, it decreases to about 0.3 in years 4 and 5 and then decreases to about 0.15 in years 6-8. At Level 2, similarly, the hazard rate of separation shows little variation over the sample periods compared to the hazard rate of promotion to Level 3, which at first increases, between the first and the second year of tenure, and then decreases, between the second and the sixth year of tenure. At Level 3 the separation hazard is roughly constant at about 0.1, but the significance of this pattern is limited by the small number of observations available.[15]

Table 3 displays, for each year since entry, the proportion of employees at Levels 1, 2 and 3 who receive a rating of 1 (high), as well as the proportion of employees at each level who receive a high rating and are assigned to the next level in the following period, i.e., the fraction of high rating among promoted employees.[16] The empty entries in the first row are due to the fact that all employees are assigned Level 1 when hired. The empty entries for workers at Level 1 promoted to Level 2 are a consequence of the fact that no employee was promoted to Level 2 after the fifth year

[15] The number of individuals employed at Level 3 in period 3 is 144, in period 3 and 4 is 129, in periods 3 to 5 is 111, in periods 3 to 6 is 100, in periods 3 to 7 is 89 and in periods 3 to 8 is 80. At high tenures, the number of retained managers at Level 3 reduces to 74, in periods 3 to 9, and to 66, in periods 3 to 10.

[16] Attention has been restricted to the sub-sample of employees with no rating information missing.

since entry. Analogously, the empty entries for promoted workers from Level 2 to Level 3 are due to the fact that no employee was promoted to Level 3 after the fifth year since entry.

**Table 2. Hazard Rates of Exit and Promotion by Level (16 or More Years of Education at Entry, Missing Ratings - 1,552 Employees)**

| Years at Level | Level 1 to Exit | Level 1 to Level 2 | Level 2 to Exit | Level 2 to Level 3 | Level 3 to Exit |
|---|---|---|---|---|---|
| 1 | 0.133 | 0.396 | 0.155 | 0.221 | 0.097 |
| 2 | 0.130 | 0.486 | 0.172 | 0.556 | 0.140 |
| 3 | 0.106 | 0.436 | 0.202 | 0.471 | 0.099 |
| 4 | 0.107 | 0.298 | 0.088 | 0.294 | 0.110 |
| 5 | 0.083 | 0.306 | 0.095 | 0.286 | 0.101 |
| 6 | 0.114 | 0.182 | 0.077 | 0.077 | 0.075 |
| 7 | 0.065 | 0.129 | 0.000 | 0.273 | 0.108 |
| 8 | 0.160 | 0.160 | 0.125 | 0.000 | - |
| 9 | 0.000 | 0.235 | - | - | - |

**Table 3. Fraction of High Ratings Among Employees at Level and Promoted Into a Level (16 or More Years of Education at Entry, No Rating Missing - 502 Employees)**

| Years Since Entry | Level 1 | Promoted to Level 2 | Level 2 | Promoted to Level 3 | Level 3 |
|---|---|---|---|---|---|
| 0 | 0.510 | 0.518 | - | - | - |
| 1 | 0.362 | 0.433 | 0.567 | 0.810 | - |
| 2 | 0.200 | 0.211 | 0.388 | 0.516 | 0.810 |
| 3 | 0.118 | 0.250 | 0.155 | 0.273 | 0.500 |
| 4 | 0.143 | 0.000 | 0.243 | 0.333 | 0.400 |
| 5 | 0.000 | 0.000 | 0.154 | 0.000 | 0.208 |
| 6 | 0.000 | - | 0.167 | - | 0.118 |
| 7 | - | - | 0.000 | - | 0.111 |
| 8 | - | - | 0.000 | - | 0.250 |
| 9 | - | - | - | - | 1.000 |

In each period the proportion of employees at Levels 1 or 2 receiving a high rating decreases over time and is significantly smaller than the fraction of promoted workers with a high rating. Moreover, the frequency of high ratings is larger among employees promoted earlier from either Level 1 to Level 2 or from Level 2 to Level 3 than among employees promoted after longer tenures. At Level 3 as well, a part from periods 8, 9 and 10, the proportion of employees receiving a high rating decreases over time.

### 3.3 Evidence from the Data and Predictions of the Model

As implied by the model, the probability of employment at the firm at any level is decreasing over time, because an increasing number of high performance realizations must occur for the firm to be willing to retain a worker. This is reflected in the data by the fact that the fraction of individuals employed at Levels 1, 2 or 3 eventually decreases over time. Moreover, the probability of being assigned to Levels 2 and 3 increases only at low levels of tenure, suggesting that employees are sorted at Level 1 according to their perceived ability, before being allocated to higher levels. The intuition from the model behind these patterns is that, at Level 2, the decrease in the probability of employment is due to the fact that good performing employees are eventually promoted to Level 3, if retained. At Level 3 it is the combined result of the selectivity of the firm's retention criterion and of the existence of an exogenous separation shock. The result that the posterior belief must be sufficiently high for a worker to be employed at Level 3, and the fact that firing a low performing worker at Level 3 can be more profitable than demoting him to Level 2 (if the change in the posterior belief after a bad performance realization is sufficiently large), together imply that workers assigned to Levels 2 and 3 might be fired and not demoted.

By comparing, from Table 3, the fraction of workers receiving in each year a high rating with the fraction of workers employed at each level, from Table 1, it follows that employees who are retained at the firm at any level, but not promoted, are those whose performance ratings is on average lower, i.e., promoted workers have highest assessed ability. This evidence is consistent with the equilibrium result that employees are progressively assigned to Levels 2 and 3, as the assessment of their talent, as revealed by their performance on the job, improves. The fact that the probability of being assigned to any level eventually decreases over time, as well as the fraction of employees receiving a high rating at each level, is also consistent with the prediction that workers whose assessed ability decreases are those more likely to leave the firm.

## 4 Empirical Analysis

### 4.1 Solution Method

Although the model does not admit a closed-form solution, it can be solved numerically for the firm's unknown value function $\Pi(\cdot)$ and the job-specific values $\Pi_k(\cdot)$, $k = 1, 2, 3$. As argued in Section 2, the value function $\Pi(\cdot)$ is a fixed point of a contraction mapping. Since the belief $\phi$ about the worker's ability being high is the only state variable in the firm's dynamic programming problem, the state space reduces to the unit interval. Therefore, $\Pi_k(\cdot)$ can be computed recursively by value function iteration. For computational reasons, the state space has been discretized in a uniform grid of 600 equidistant points on the interval [0,1].

When the distribution of revenue realizations is not symmetric across types, i.e., the probability of a high rating for a high ability worker is different from the probability of a low rating for a low ability worker (i.e., $\alpha_k \neq 1 - \beta_k$ for some $k = 1, 2, 3$, where $\alpha_k$ is the probability of a high rating at job $k$ for a worker of high ability and $\beta_k$ for a worker of low ability), the process of posterior beliefs visits different states along each equilibrium sample path, for given prior belief $\phi_1$. This implies that, for every belief value on the grid, the updated posterior computed by Bayes' rule can be a point outside the grid. Note that this problem would also arise in the symmetric case, as long as the probability of a high rating was different across tasks.[17] To ensure that the updated posterior from each possible belief value on the grid is itself a grid point, a nearest neighborhood procedure has been adopted, to select the value on the grid closest to the exact Bayes' update.

The firm's optimal employment (i.e., retention and task allocation) policy is then computed by determining, for each belief value on the grid, the task which generates the highest expected discounted profit, by direct comparison of the alternative-specific values, as computed from $\Pi(\cdot)$.

## 4.2 Estimation Method

Given that in the model the firm and the worker are assumed to be endowed with the common prior $(1 - \phi_1, \phi_1)$ over the ability space $\{\underline{\theta}, \overline{\theta}\}$ at the beginning of period 1, the distribution of prior beliefs is not determined by the model. In estimation we assume that the probability $\phi_1$ that the worker is of high ability is drawn from a beta distribution over the set of belief values for which the assignment of Level 1 is profitable for the firm.[18] Denote the vector of structural parameters to be estimated by

$$\psi = (a_\beta, b_\beta, \delta, (\alpha_k, \beta_k, \overline{y}_k, \underline{y}_k, \xi_k, E_k(\overline{\theta}), E_k(\underline{\theta}))_{k=1}^3),$$

where $a_\beta$ and $b_\beta$ are the parameters of the beta distribution from which the initial prior $\phi_1$ is drawn, $\delta$ indicates the firm and the worker's discount factor, $\alpha_k$ (for a high ability worker) and $\beta_k$ (for a low ability worker) are, for each Level $k = 1, 2, 3$, the locational parameters of the Bernoulli distribution governing output realizations, which can be high, $\overline{y}_k$, or low $\underline{y}_k$, and $\xi_k$ is the exogenous probability that the worker leaves the firm at the end of a period when assigned to Level $k$. Performance outcomes are assumed to be measured with error. The classification error rate, $E_k$, depends on the job level the worker is assigned to in a period and on the worker's true ability.[19]

Observe that the firm's reservation profit, $\Pi$, and each worker's reservation utility, $U$, act in the model as scale parameters of the expected one period return at each level, $y_k(\phi)$. As such, they

[17] Only if $\alpha_k = 1 - \beta_k$ and $\alpha_k = \alpha_{k'}$, for $k, k' = 1, 2, 3$, the non-linearity of the Bayes map could be accommodated by selecting a different belief grid for each $\phi_1$.

[18] The choice of the beta specification is motivated by its flexibility and the fact that it has a compact support, so that, in particular, $\phi_1$ can be restricted to belong to the interval of belief values $[\phi_1^*, \phi_2^*)$. In fact, for the relevant set of parameter values selected during estimation, the resulting firm's optimal employment policy is the one predicted by the model, i.e., the interval belief strategy prescribing that the worker be assigned to job 1 as long as $\phi$ lies in $[\phi_1^*, \phi_2^*)$.

[19] The interpretation of the classification error as type dependent follows the modelling hypothesis that, on average, high ability employees generate more high ratings than low ability employees. A flexible error structure allows therefore the model to fully capture differences in the probability of success across types as can be estimated from the histories of observed ratings.

cannot be separately identified from $\overline{y}_k$ and $\underline{y}_k$. For given $\alpha_k$ and $\beta_k$, in fact, a proportional change in $\overline{\Pi}$ and in $\overline{y}_k$ and $\underline{y}_k$ leaves the relative worth of the jobs in static terms unchanged. In particular, even if the one-period revenue $\Pi$ from terminating the worker and the expected one period revenue $y_k(\phi)$ from Level 1, 2 or 3 increase, the firm is indifferent between any two of the employment alternatives for the same belief values. Similarly, for given $\phi$, the same proportional increase in $U$ and decrease in $\overline{y}_k$ and $\underline{y}_k$, for all $k$, leaves $y_k(\phi)$ unchanged. Therefore, $\Pi$ and $U$ are normalized to zero.[20]

The model is estimated by smooth simulated maximum likelihood. At any time $t$ denote the vector of observed outcomes for individual $i$ by $O_{it} = (L^o_{it}, R^o_{it})$, the job level the individual is assigned to in period $t$, $L^o_{it}$, and the performance realization recorded for the period, $R^o_{it}$. Let $\theta_1 \equiv \underline{\theta}$ denote the low level of ability and $\theta_2 \equiv \overline{\theta}$ the high level. Let $s_1 \equiv e_1 \geq 16$ indicate the number of years of education of an employee at entry. The likelihood function for a sample of $N$ individuals, observed from period $t = 1$ to period $t = 10$, is given by the product over all individuals of the 10 period outcome histories of observed levels and performance ratings, conditional on their education at entry,

$$\mathcal{L}(\psi \mid s_1) = \prod_{i=1}^{N} \int_{\phi_1} \sum_{k=1}^{2} \Pr(\theta_k \mid \phi_1, s_1) \Pr(O_{i1}, ..., O_{i10} \mid \theta_k, \phi_1, s_1) dF(\phi_1 \mid s_1).$$

Since the firm and the worker' initial prior distribution over the worker's ability is not observed, the probability of each individual history has to be integrated over all possible priors. In estimation the beta distribution, which parameterizes the set of potential prior distributions, is discretized in $J$ points over the interval $[\phi^*_1, \phi^*_2)$, so that the likelihood function is approximated as

$$\mathcal{L}(\psi \mid s_1) \simeq \prod_{i=1}^{N} \sum_{j=1}^{J} \Pr(\phi_1 = \phi^j_1 \mid s_1) \sum_{k=1}^{2} \Pr(\theta_k \mid \phi^j_1, s_1) \Pr(O_{i1}, ..., O_{i10} \mid \theta_k, \phi^j_1, s_1) \tag{2}$$

where $\Pr(\theta_1|\phi^j_1, s_1) \equiv 1 - \phi^j_1$ and $\Pr(\theta_2|\phi^j_1, s_1) \equiv \phi^j_1$, $j = 1, ..., J$. In expression (2), $\Pr(\phi_1 = \phi^j_1|s_1)$ represents the probability that the firm and worker $i$'s prior belief about the worker's ability being high is $\phi^j_1$ at the beginning of period 1. Given that an individual can be either of high or of low ability with probability $\Pr(\theta_k|\phi^j_1, s_1)$ at entry, the likelihood function is obtained as the product over all individuals of the probabilities of the type-dependent outcome histories $\Pr(O_{i1}, ..., O_{i10}|\theta_k, \phi^j_1, s_1)$. The mixture over types is obtained by integrating over the prior distribution $(\phi^j_1, 1 - \phi^j_1)$. For

[20] The estimation of a version of the model which encompasses both promotion and wage dynamics is currently being implemented. In this formulation the worker's human capital is assumed to be perfectly transferable across firms. Theoretical results for the general case, in which ability can be general or firm specific, are derived in the companion paper. Following the characterization of the equilibria of interest, the specification of the wage paid, when the belief about individual $i$'s ability being high is $\phi_t$, is $\ln w^o_{ikt} = \ln a_k + \ln w_k(\phi_t) + \varepsilon_{ikt}$, if individual $i$ is assigned to Level $k = 1, 2, 3$ in period $t$. In this expression $a_k$ accounts for differences in the wage offers of different firms, as well as for possibly unobserved bonus payments, while $\varepsilon_{ikt}$ is the draw of the measurement error on wages, assumed to be normally distributed with mean zero and variance $\sigma^2_k$ at Level $k$. This formulation then allows an indirect test of the hypothesis that wages at the managerial level are set competitively. Details can be provided upon request.

each individual, the probability of his observed employment history at the firm, conditional on his education, is finally computed by weighting the prior-dependent history with the probability of a particular prior being the initial belief the firm and the individual are endowed with.

For any individual the probability of a period-$t$ outcome pair can be factored in the product of the probability of the assigned level and of the performance signal observed in the period, conditional on this level. The conditional probability of an individual $i$'s outcome history can therefore be expressed as

$$\begin{aligned}\Pr(O_{i1}, ..., O_{i10} \mid \theta_k, \phi_1^j, s_1) \;=\; & \Pr(L_{i1}^o \mid \theta_k, \phi_1^j, s_1) \Pr(R_{i1}^o \mid \theta_k, L_{i1}^o) \cdots \\ & \cdot \Pr(L_{i10}^o \mid \theta_k, \phi_1^j, R_{i1}, ..., R_{i9}, s_1) \cdot \Pr(R_{i10}^o \mid \theta_k, L_{i10}^o) \end{aligned} \tag{3}$$

where $L_{it}^o \in \{L_0, L_1, L_2, L_3\}$ indicates the level assignment, with $L_0$ representing no employment, and $R_{ij} \in \{\emptyset, 0, 1\}$, $j = 1, ..., 10$ the actual performance outcome realized in period $t$ (note that the rating of a worker who has left the firm is missing by construction). The probability of each observed level is computed conditional on the worker's unobserved ability (which determines the probability distribution of the true performance signal at each task), the initial prior $\phi_1$ (which determines the probability of the worker's initial job assignment at the firm) and the sequence of past realized ratings (which, together with the initial prior, determine the value of the updated posterior). The probability of the observed rating, instead, only depends upon the worker's actual ability and the level assigned, from our assumption that the distribution of revenue realizations at Level $k = 1, 2, 3$ is bernoulli with parameter $\alpha_k$, if the worker is of high ability, and $\beta_k$, if the worker is of low ability.

In expression (3) it is implicit that, given the bernoulli process governing output realizations at each level, at any time $t$ the initial prior and the sequence of *true* performance realizations are sufficient statistics for the updated posterior. Specifically, $\Pr(L_{it}^o|\theta_k, \phi_1^j, R_{i1}, ..., R_{it-1}, s_1) = \Pr(L_{it}^o|\theta_k, \phi_t^j, s_1)$, where $\phi_t^j$ represents the updated posterior at the beginning of period $t$ from the prior $\phi_1^j$ and the sequence of actual performance outcomes from period 1 through $t-1$, $(R_{i1}, ..., R_{it-1})$.

For each individual, the probability of the per-period level $L_{it}^o = L_r$ is calculated as $\Pr(\Pi_r(\phi) = \max\{\Pi_0(\phi), \Pi_1(\phi), \Pi_2(\phi), \Pi_3(\phi)\})$, viewed, for the purpose of estimation, as a function of the parameters of the model conditional on the data, and it is computed by a kernel smoothed frequency simulator. Specifically, the probability of the observed level in each period, for given initial prior $\phi_1^j$, is simulated over $S$ possible realizations of the performance rating in the period and smoothed through a logistic kernel with bandwidth parameter $\tau$.[21] The corresponding kernel is computed as

$$\begin{aligned}\Pr(L_{it}^o = L_r \mid \theta_k, \phi_t^{js}, s_1) \;=\; & \exp\left[\frac{\Pi_r(\phi_t^{js}(\theta_k)) - \max_l\{\Pi_l(\phi_t^{js}(\theta_k))\}}{\tau}\right] \\ & \cdot \left\{\sum_{m=0}^{3} \exp\left[\frac{\Pi_m(\phi_t^{js}(\theta_k)) - \max_l\{\Pi_l(\phi_t^{js}(\theta_k))\}}{\tau}\right]\right\}^{-1}\end{aligned}$$

[21] In the actual estimation, the bandwidth has been set to 10, based on sensitivity analysis. The procedure is an application of the measurement error technique introduced by McFadden [1989]. See also Keane and Wolpin [1997] and Eeckstein and Wolpin [1999].

at the $j$-th draw of the initial prior and the $s$-th simulation draw of the performance realization, with $L_r \in \{L_0, L_1, L_2, L_3\}$. In the expression $\phi_t^{js}(\theta_k)$ denotes the updated posterior from the prior $\phi_1^j$ and the sequence of performance ratings $(R_{i1s}, ..., R_{it-1s})$ from period 1 to period $t-1$, simulated conditional on the worker's true ability, i.e., $\phi_t^{js}(\theta_k) = \varphi(\phi_1^j \mid R_{i1s}(\theta_k), ..., R_{it-1s}(\theta_k))$.[22] The probability of the observed level is then computed as the average of the above kernel over the $S$ simulations of the performance rating,

$$\Pr(L_{it}^o = L_r \mid \theta_k, \phi_t^j, s_1) \simeq \sum_{s=1}^{S} \frac{\Pr(L_{it}^o = L_r \mid \theta_k, \phi_t^{js}, s_1)}{S}.$$

As mentioned, to avoid zero-probability events contributing to the likelihood function, and given the inherent noisiness of the performance appraisal process, it is assumed that performance ratings are measured with error. Formally, the conditional probability of observing a rating $R_{it}^o \in \{0,1\}$ in period $t$ at level $L_{it}^o = L_k \in \{L_1, L_2, L_3\}$, if the true performance is $R_{it}$ and the worker's ability $\theta_k \in \{\underline{\theta}, \overline{\theta}\}$, is given by

$$\begin{aligned}
\Pr(R_{it}^o = 1 \mid R_{it} = 1, \theta_k, L_{it}^o) &= E_k(\theta_k) + (1 - E_k(\theta_k)) \Pr(R_{it} = 1 \mid \theta_k, L_{it}^o) \\
\Pr(R_{it}^o = 1 \mid R_{it} = 0, \theta_k, L_{it}^o) &= (1 - E_k(\theta_k)) \Pr(R_{it} = 1 \mid \theta_k, L_{it}^o)
\end{aligned}$$

where $\Pr(R_{it} = 1 | \overline{\theta}, L_{it}^o = L_k) = \alpha_k$ and $\Pr(R_{it} = 1 | \underline{\theta}, L_{it}^o = L_k) = \beta_k$, $k = 1, 2, 3$.[23] In this way the model of misclassification is characterized by four rates, out of all the possible combinations of observed and true choices, since

$$\begin{aligned}
\Pr(R_{it}^o = 1 \mid \theta_k, L_{it}^o) = &\Pr(R_{it}^o = 1 \mid R_{it} = 1, \theta_k, L_{it}^o) \Pr(R_{it} = 1 \mid \theta_k, L_{it}^o) \\
&+ \Pr(R_{it}^o = 1 \mid R_{it} = 0, \theta_k, L_{it}^o) \Pr(R_{it} = 0 \mid \theta_k, L_{it}^o).
\end{aligned}$$

In this specification, the classification error is unbiased: the (conditional) probability of observing a high output realization is the same as the (conditional) probability that a good output truly occurs, i.e., $\Pr(R_{it}^o = 1 | \theta_k, L_{it}^o) = \Pr(R_{it} = 1 | \theta_k, L_{it}^o)$. Unbiasedness implies that the classification rates are linear in the true choice probabilities. As the probability of the true choice converges to one, the probability of the observed choice converges to one as well, i.e., unbiasedness is preserved in the limit since the probability of a correct classification increases linearly from $E_k(\theta_k)$ to one as the true choice probability approaches one. In other words, as $\Pr(R_{it} = 1 | \theta_k, L_{it}^o) \to 1$, $\Pr(R_{it}^o = 1 | R_{it} = 1, \theta_k, L_{it}^o) \to 1$. In addition, when the probability of the true choice goes to zero, $E_k(\theta_k)$

[22] Notice that the number of performance ratings simulated in each period is constant across individuals and prior draws.

[23] See Keane and Wolpin [2001] and Keane and Sauer [2003]. When simulating outcomes for given parameter values, the sequence of reported choices with errors is constructed by drawing a sequence $\{U_{it}\}_{t=1}^T$ of $T = 10$ deviates from a uniform number generator for each individual $i$ and comparing these draws with the classification error rates. The comparison determines whether choices are correctly reported, by the following rule: given $R_{it} = 1$, if $U_{it} < \Pr(R_{it}^0 = 1 | R_{it} = 1, \theta_k, L_{it}^o)$, then $R_{it}^0 = 1$, and $R_{it}^0 = 0$, otherwise. Similarly, given $R_{it} = 0$, if $U_{it} < \Pr(R_{it}^0 = 0 | R_{it} = 0, \theta_k, L_{it}^o)$, then $R_{it}^0 = 0$, and $R_{it}^0 = 1$, otherwise.

approximates the conditional probability of observing the true choice, since $\Pr(R_{it} = 1|\theta_k, L_{it}^o) \to 0$ implies $\Pr(R_{it}^o = 1|R_{it} = 1, \theta_k, L_{it}^o) \to E_k(\theta_k)$. In this sense $E_k(\theta_k)$ can be interpreted as a base classification error rate. In estimation, $E_k(\theta_k)$ is treated as a free parameter and it is the only parameter on which unbiasedness depends.

Given the assumed error structure for the performance signal, the associated probability of the observed rating is computed as

$$\Pr(R_{it}^o = R_r \mid \theta_k, L_{it}^o) \simeq \sum_{s=1}^{S} \frac{\Pr(R_{it}^o = R_r \mid R_{its} = R_{rs}, \theta_k, L_{it}^o)\Pr(R_{its} = R_{rs} \mid \theta_k, L_{it}^o)}{S}$$

where $R_{rs}$ denotes the realization of the performance signal at the $s$-th simulation, with $R_r, R_{rs} \in \{\emptyset, 0, 1\}$. The sequence $(R_{it1}, ..., R_{itS})$ of period $t$ simulation draws is then used to compute the vector of period-$t+1$ updated posteriors $(\phi_{t+1}^{j1}, ..., \phi_{t+1}^{jS})$.

Notice that the entire set of the model parameters enters the likelihood through the choice probabilities, which are computed from solving the firm's dynamic programming problem. The maximization of the likelihood function involves an iterative process between the numerical solution of the firm's dynamic programming problem, for given parameter values, and the computation of the likelihood function.[24]

# 5 Estimation Results

## 5.1 Parameter Estimates

The qualitative implication of the model that experimenting on a worker's unobserved ability is an important determinant of job to job transitions inside a firm is confirmed by preliminary estimation results.[25] Table 4 reports the value of the vector $\psi = (a_\beta, b_\beta, \delta, (\alpha_k, \beta_k, \underline{y}_k, \overline{y}_k, \xi_k, E_k(\underline{\theta}), E_k(\overline{\theta}))_{k=1}^3)$ of structural parameters, estimated from the sample of 502 managers entering the firm at Level 1 between 1970 and 1979, with at least 16 years of education and no level assignment or performance rating missing. Relevant descriptive statistics for the sample, together with standard errors for the parameter estimates, are reported in Appendix B.[26]

[24] Observe that, given our simulation technique, as long as the number of simulations, as compared to the number of individuals, grows arbitrarily large, the simulated maximum likelihood estimates are consistent and asymptotically normal. If this ratio is bounded away from infinity, the estimates are still consistent, but the limiting distribution is normal with mean not equal to zero, i.e., there is a bias.

[25] Estimation of the model from the sample of 1,552 managers entering the firm at Level 1 between 1970 and 1979, with at least 16 years of education at entry and no level information missing, is currently pursued. The probability that a rating is missing for each type of worker is then estimated as an additional structural parameter. Details can be provided upon request.

[26] In the present version of the estimation, separation rates at Levels 1 ($\xi_1$) and 2 ($\xi_2$) have been set to zero. The asymptotic variance of the estimates is computed using the so-called BHHH estimator, proposed by Berndt, Hall, Hall and Hausmal in 1974 and based on the outer product of the gradient of the log-likelihood at the estimated parameter values.

**Table 4. Parameter Estimates**

| | | | |
|---|---|---|---|
| $a_\beta$ | 1.000 | $\underline{y}_1$ | $-2,446.885$ |
| $b_\beta$ | 1.000 | $\underline{y}_2$ | $-5,986.493$ |
| $\alpha_1$ | 0.869 | $\underline{y}_3$ | $-880,226.430$ |
| $\alpha_2$ | 0.778 | $\xi_1$ | - |
| $\alpha_3$ | 0.999 | $\xi_2$ | - |
| $\beta_1$ | 0.069 | $\xi_3$ | 0.564 |
| $\beta_2$ | 0.000 | $E_1(\overline{\theta})$ | 0.010 |
| $\beta_3$ | 0.700 | $E_2(\overline{\theta})$ | 0.002 |
| $\delta$ | 0.950 | $E_3(\overline{\theta})$ | 0.111 |
| $\overline{y}_1$ | 50.940 | $E_1(\underline{\theta})$ | 0.000 |
| $\overline{y}_2$ | $3,599.936$ | $E_2(\underline{\theta})$ | 0.987 |
| $\overline{y}_3$ | $40,846.745$ | $E_3(\underline{\theta})$ | 0.001 |

From these estimated values, as predicted by the model the firm's optimal employment policy is an interval belief strategy, which prescribes that the worker be assigned to Level 1 if $\phi \in [0.052, 0.503)$, to Level 2 if $\phi \in [0.503, 0.993)$ and to Level 3 if $\phi \in [0.993, 1]$, but that he be not employed if $\phi \in [0, 0.052)$. A number of theoretical restrictions under which this policy, characterized in Section 2, is the firm's optimal employment policy are also satisfied. Notice first that the distribution of output signals at the three levels is asymmetric across types, i.e., $\alpha_k \neq \beta_k$ for $k = 1, 2, 3$. Moreover, at each job $k = 1, 2, 3$, the distribution of output realizations when the worker is of high ability first-order stochastically dominates the corresponding distribution when he is of low ability, i.e., $\alpha_k > \beta_k$. This implies, as posited by the model, that observing a high rating improves the firm's assessment that the worker is of high ability. Given these values for $\alpha_k$ and $\beta_k$, the estimated size of the output realizations at the three jobs, $\underline{y}_k$ and $\overline{y}_k$, satisfies assumptions (A1)-(A2), i.e., $y_3(\overline{\theta}) > y_2(\overline{\theta}) > y_1(\overline{\theta})$ and $y_1(\underline{\theta}) > y_2(\underline{\theta}) > y_3(\underline{\theta})$, with $y_1(\underline{\theta}) = -2,274.612$, $y_2(\underline{\theta}) = -5,986.075$ and $y_3(\underline{\theta}) = -235,425.955$.

In particular, from the fact that $\Pi = 0 > y_1(\overline{\theta}) = -276.456$, while $y_2(\overline{\theta}) = 1472.475$ and $y_3(\overline{\theta}) = 40,842.563$, it follows that the expected continuation profit from assigning a worker to Level 1 is sufficiently large to compensate the one-period profit loss from employing him. The fact that, at the estimated parameter values, the expected continuation value at Level 2 exceeds the one at Level 1 at each belief also implies that the (gross) informational value at Level 2 is larger than the one at Level 1. The maximal difference is of the order of 1,360, when $y_2(\phi) = 1,472.4$. Finally, a value of $\delta = 0.95$ is consistent with the yearly observations used in estimation, given that it implies an annual interest rate on a risk free asset of 4.75 percent.

## 5.2 Within-Sample Fit

We will now present evidence of the model's within-sample fit by looking at the distribution of managerial employees across Levels 1, 2 and 3, over the first nine years after entry. The observed and predicted fraction of those managerial workers, entering the firm at Level 1, who are assigned to Levels 1, 2 and 3 or leave the firm, in each of the nine years after entry, are reported in Table 5.

**Table 5**. **Proportion of Employees at Levels 1 and 2, Observed (BGH) and Predicted (DP) (16 or More Years of Education at Entry - 502 Employees)**

| **Years Since Entry** | **Level 1 (BGH)** | **Level 1 (DP)** | **Level 2 (BGH)** | **Level 2 (DP)** | **Level 3 (BGH)** | **Level 3 (DP)** | **Exit (BGH)** | **Exit (DP)** |
|---|---|---|---|---|---|---|---|---|
| 0 | 1.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0.347 | 0.330 | 0.327 | 0.262 | 0.000 | 0.000 | 0.327 | 0.408 |
| 2 | 0.100 | 0.058 | 0.277 | 0.087 | 0.042 | 0.168 | 0.582 | 0.686 |
| 3 | 0.034 | 0.071 | 0.141 | 0.021 | 0.084 | 0.115 | 0.741 | 0.793 |
| 4 | 0.014 | 0.010 | 0.074 | 0.020 | 0.070 | 0.064 | 0.843 | 0.906 |
| 5 | 0.008 | 0.009 | 0.026 | 0.003 | 0.048 | 0.039 | 0.918 | 0.949 |
| 6 | 0.002 | 0.004 | 0.012 | 0.003 | 0.034 | 0.019 | 0.952 | 0.974 |
| 7 | 0.000 | 0.001 | 0.006 | 0.001 | 0.018 | 0.010 | 0.976 | 0.988 |
| 8 | 0.000 | 0.001 | 0.002 | 0.000 | 0.008 | 0.005 | 0.990 | 0.994 |
| 9 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 | 0.002 | 0.998 | 0.997 |

As it can be seen, the model succeeds in capturing the dynamic profile of the probability of continuous assignment to Level 1, which is steeply decreasing for the sample of observed employees over the ten year period. The pattern of assignment to Level 2 implied by the model also shares the same qualitative features of the profile observed in the data: it sharply increases the second year after entry and then decreases throughout. The greatest discrepancy between the observed and predicted fraction of managers employed at Level 2 is in the second and third year since entry, when the fraction of employees in the data assigned to Level 2 is significantly greater than the proportion predicted by the model, with a difference, respectively, of 0.19 and 0.12.[27]

As for the observed and predicted fraction of employees who are assigned to Level 3, the hump-shaped pattern observed in the data, increasing in the first three years since entry and then decreasing, is successfully captured by the model. Nonetheless, the proportion of employees assigned to Level 3 in the second and third period since entry, simulated from the model, is substantially larger than the fraction observed in the data. This is a consequence of the fact that the model predicts a smaller proportion of employees at Level 2 in those same periods than the one actually observed, while the observed and predicted exit rate, as well as the observed and predicted proportion of employees at

[27]One of the dimensions along which the model fit could be improved is by allowing for the presence of unobserved heterogeneity in the probability of a success at Level 2, for each worker type. This modification would in fact introduce a more flexible parametrization of the determinants of profit at Level 2, which could allow the model to match more closely the pattern of allocation to Level 2 and then to Level 3 at high tenures.

Level 1, are fairly similar in those years. In fact, the fraction of employees leaving the firm predicted by the model is close to the fraction observed in the data, with the largest difference of 0.104 in the second period after entry.

## 6 The Value of Information

In general there may be no obvious units to measure the amount of information available to a decision maker. The question is nonetheless meaningful in the context of a broader decision problem, which involves choosing an information structure. When a worker is assigned to a job position, the revenue generated in a period is not only the source of the firm profit, but it provides the firm with additional information about the worker's ability, given that the likelihood of observing either a high or a low rating (proxy for high or low revenue) depends on the worker's underlying ability. Specifically, the choice of a job affects the distribution of the performance signals generated in a period and therefore the distribution of the firm's posterior beliefs. In a sense, then, choosing to which job to assign the worker can be viewed as choosing which information to generate about his ability, i.e., which experiment to perform in order to learn about his unobserved human capital.

Notice that, if the firm did not observe the revenue produced by the worker on the job in a period, it would not be able to condition its future assignment decisions on it. In this case, the expected discounted profit from assigning the worker to job $k$ would be

$$\Pi_k(\phi) = y_k(\phi) - U + \delta(1-\xi_k)\Pi_k(\phi) + \delta\xi_k\overline{\Pi}$$

so that $\Pi_k(\phi) = y_k(\phi) - U + \delta\xi_k\overline{\Pi}/[1-\delta(1-\xi_k)]$.[28] However, since the firm can condition its decision of which job to assign the worker in period $t+1$ on the performance signal observed in period $t$, a natural measure of the (gross) value of information is the extra expected profit, from period $t+1$ on, from choosing task $k$ over the task which maximizes the expected period profit, task $s$. This value can then be quantified as the difference between the firm's maximal expected continuation profit, function of its current period choice of job $k$, $E_k[\Pi(\tilde{\phi})|\phi]$, minus $y_s(\phi) - U + \delta\xi_s\overline{\Pi}/[1-\delta(1-\xi_s)]$.[29] Analogously, the firm's willingness to pay for the maximal amount of information, from assigning the worker to each job $k$, can be measured as the difference, at each belief, between $(i)$ the expected continuation value from the most informative experiment at job $k$, which would immediately reveal the worker's ability after one period, and $(ii)$ the expected continuation value from assigning the worker to job $k$.

As discussed the firm incurs an opportunity cost in generating information about a worker's ability. The option value of this information is the expected one-period profit loss the firm incurs by

[28] If payoffs were normalized by $1-\delta$, so as to be expressed as per period averages, and the separation rate at each job was zero, this value would be the same as the firm's profit from the static game, i.e., on average the firm would receive his period profit.

[29] See this approach in Chade and Schlee [2002] for a discussion.

choosing to assign the worker to job $k$ rather than to the most profitable job for that period. Then, the above discussion implies that the net value of information to the firm can be measured as

$$\begin{aligned} I_k(\phi) &\equiv \delta\left\{(1-\xi_k)E_k[\Pi(\tilde{\phi})|\phi] + \delta\xi_k\overline{\Pi} - \frac{y_s(\phi) - U + \delta\xi_s\overline{\Pi}}{1-\delta(1-\xi_s)}\right\} - [y_s(\phi) - y_k(\phi)] \\ &= y_k(\phi) - U + \delta(1-\xi_k)E_k[\Pi(\tilde{\phi})|\phi] + \delta\xi_k\overline{\Pi} - \frac{(1-\delta\xi_s)[y_s(\phi) - U] + \delta^2\xi_s\overline{\Pi}}{1-\delta(1-\xi_s)} \end{aligned}$$

that is, approximately the extra payoff from the dynamic game over obtaining perpetually the static game profit. Observe that, given that the firm's value function is convex in the posterior belief, the firm always values new information as long as there is uncertainty about the worker's ability, in the sense that it always prefers a riskier distribution of posterior beliefs to a less risky one, as discussed in Section 2.

The objective of the present Section is to assess the impact on the net value of information to the firm and on learning (measured as the probability of retaining a high ability worker) of altering specific parameters of the environment from their estimated values. The effect of changes in the structural parameters of the model on these quantities are in principle not obvious. Modifications of some parameters, namely $\alpha_k$ and $\beta_k$, affect directly the informational content of job $k$. Nevertheless, all the structural parameters $\delta$, $\alpha_k$, $\beta_k$, $\underline{y}_k$, $\overline{y}_k$ and $\xi_k$, $k = 1, 2, 3$, have an impact on the firm's own *valuation* of information, since they affect the degree of convexity of $\Pi(\cdot)$, and, in this way, the value of information.

### 6.1 Changes in the Value of Information and Incomplete Learning

One of the purposes of the estimation exercise is to determine the value of endogenous information acquisition to the firm, and to quantitatively evaluate the changes in this value under alternative scenarios, simulated from the benchmark case, in which parameters are fixed at their estimated values. Estimates of the parameters of the model also allow to uncover the equilibrium relationship between the value of information and the amount of learning which takes place through employment, measured as the probability of retaining a high ability worker at either Level 1, Level 2 or Level 3 in each period. An important dimension along which counterfactual experiments are of interest is therefore in assessing the impact of changes in the value of information on the (inefficient) turnover of high ability workers.

Understanding the effect of learning on firm-level allocation decisions has significant implications for the labor market experience of workers and for a firm's incentive to employ them. It also enables us to make predictions about the effectiveness of policies that aim at improving monitoring of performance, i.e., the quality of the information generated through employment, at each level of the firm's hierarchy. The counterfactual evaluations we will perform aim specifically at investigating: ($i$) the impact on the value of information acquisition to the firm, and ($ii$) the resulting comparative dynamic effect on workers' career prospects, of:

(1) changes in the firm's degree of time impatience, which parameterizes the firm's incentive to experiment on ability, to $\delta = 0.50$ and $\delta = 0.99$;

(2) an increase in the precision of prior information, i.e., a reduction in the dispersion of the distribution of possible prior beliefs about the worker's unobserved ability, to $a_\beta = a_\beta = 50$;

(3) an increase in the accuracy of the firm's monitoring technology, i.e., the probability of a high rating for a worker of either ability at Levels 1 and 2;

(4) a reduction to zero in the size of the output realizations and in the probability of success for each type of worker at Level 2, i.e., the case in which only Level 1, the entry job, and Level 3, the statically most profitable job, are available.

### 6.2 Experiment 1: Different Degrees of Time Patience

In order to illustrate the value of information acquisition implied by the parameter estimates, we compare the model's prediction on the distribution of employees across the three managerial levels, together with the fraction of employees leaving the firm, with the predictions from a model in which $\delta = 0.50$, i.e., intermediate degree of time impatience, and a model in which $\delta = 0.99$, i.e., close to maximal time patience. Tables 6 and 7 report the predicted fraction of employees at each level and leaving the firm in each year since entry, for the benchmark case and the simulated scenarios.

**Table 6. Predicted Fraction of Employees at Levels 1 and 2**

| Years Since Entry | Level 1 ($\delta = 0.95$) | Level 1 ($\delta = 0.50$) | Level 1 ($\delta = 0.99$) | Level 2 ($\delta = 0.95$) | Level 2 ($\delta = 0.50$) | Level 2 ($\delta = 0.99$) |
|---|---|---|---|---|---|---|
| 0 | 1.000 | 1.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0.330 | 0.000 | 0.525 | 0.262 | 0.241 | 0.314 |
| 2 | 0.058 | 0.000 | 0.174 | 0.087 | 0.000 | 0.291 |
| 3 | 0.071 | 0.000 | 0.093 | 0.021 | 0.000 | 0.309 |
| 4 | 0.010 | 0.000 | 0.046 | 0.020 | 0.000 | 0.322 |
| 5 | 0.009 | 0.000 | 0.019 | 0.003 | 0.000 | 0.319 |
| 6 | 0.004 | 0.000 | 0.010 | 0.003 | 0.000 | 0.324 |
| 7 | 0.001 | 0.000 | 0.005 | 0.001 | 0.000 | 0.324 |
| 8 | 0.001 | 0.000 | 0.002 | 0.000 | 0.000 | 0.324 |
| 9 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.324 |

As expected, when the discount factor is $\delta = 0.50$, the firm's willingness to employ workers decreases, since the value of current information for the profitability of future assignment decisions is smaller. Indeed, the range of beliefs for which employment is profitable decreases, i.e., the lowest belief for which the firm is willing to employ a worker at Level 1 is approximately $\phi = 0.20$, as compared to $\phi = 0.05$ in the benchmark case. The firm's degree of time impatience has a significant effect on the pattern of exit as well: for the same rate of exogenous separations, the fraction of managers who will leave the firm after the first period almost doubles, from 0.481 in the benchmark case to 0.876.

**Table 7. Predicted Fraction of Employees at Level 3 and Leaving the Firm**

| Years Since Entry | Level 3 ($\delta = 0.95$) | Level 3 ($\delta = 0.50$) | Level 3 ($\delta = 0.99$) | Exit ($\delta = 0.95$) | Exit ($\delta = 0.50$) | Exit ($\delta = 0.99$) |
|---|---|---|---|---|---|---|
| 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0.000 | 0.000 | 0.000 | 0.408 | 0.759 | 0.162 |
| 2 | 0.168 | 0.145 | 0.000 | 0.686 | 0.759 | 0.534 |
| 3 | 0.115 | 0.095 | 0.000 | 0.793 | 0.905 | 0.598 |
| 4 | 0.064 | 0.041 | 0.000 | 0.906 | 0.959 | 0.631 |
| 5 | 0.039 | 0.018 | 0.000 | 0.949 | 0.982 | 0.662 |
| 6 | 0.019 | 0.008 | 0.000 | 0.974 | 0.992 | 0.666 |
| 7 | 0.010 | 0.004 | 0.000 | 0.988 | 0.996 | 0.672 |
| 8 | 0.005 | 0.001 | 0.000 | 0.994 | 0.999 | 0.674 |
| 9 | 0.002 | 0.001 | 0.000 | 0.997 | 0.999 | 0.674 |

When $\delta = 0.99$, the cut-off belief which makes the firm willing to assign the worker to Level 3 is almost one. This implies that most of the workers, whose ability is sufficiently high for being retained at the firm, are assigned to Level 2 for a longer period of time. In particular, in the first ten periods of employment none of them is assigned to Level 3. This has also a clear impact on the fraction of employees leaving the firm. The fact that, when a worker is of high skill, output realizations are noisier signals of ability at Level 2 than at Level 3 (i.e., $\alpha_2 < \alpha_3$), together with the fact that Level 2 is profitable for higher belief values than in the benchmark case, imply that employees at Level 2 are more likely to be retained rather than fired, for the same sequence of observed ratings. This is reflected in the smaller fraction of employees leaving the firm after the third period since entry.

When $\delta = 0.50$, as compared to the case in which $\delta = 0.95$, the value of information is higher than in the benchmark case for any $\phi \geq 0.26$ and the increase can be as large as of the order of 444 percent. This result is due to the fact that, even if in principle the firm values information more when it is less time impatient, given that it attaches a greater weight to his future expected profit, the fact that exogenous separations at Level 3 occur with high probability depresses significantly the firm's expected discounted profit from assigning an employee to any job. As expected, when $\delta = 0.99$ the firm has nevertheless a stronger incentive to employ a worker to learn about his ability. In fact, employment starts being profitable for the firm when $\phi = 0.002$, where the firm's expected discounted profit increases by as much as 400 percent. Otherwise, the change in the value of information, compared to the benchmark case, ranges from approximately 0 percent, when $\phi = 0.075$, to -60 percent, when $\phi = 0.85$.[30]

[30]However, the difference in the values of information is non monotonic at high belief values.

### 6.3 Experiment 2: Increased Precision of Prior Information

Recall that a beta distribution with parameters $a_\beta = b_\beta = 1$ parameterizes the set of prior distributions for the firm and a worker over the worker's true ability. The variance of a beta distribution with parameters $a_\beta$ and $b_\beta$ is equal to $a_\beta b_\beta/((a_\beta + b_\beta)^2(a_\beta + b_\beta + 1))$, so that, when $a_\beta = b_\beta = 50$ as compared to when $a_\beta = b_\beta = 1$, it decreases from 0.083 to 0.002. An increase in $a_\beta$ and $b_\beta$ is then equivalent to a reduction in the dispersion of prior beliefs about a newly hired worker, still consistent with the worker being assigned to Level 1 at entry. Given the estimated values of the parameters, however, a reduction of $5,000$ percent in this dispersion, has no significant impact on the dynamics of job assignment inside the firm, a part for a decrease in the fraction of workers leaving the firm in the second period after entry.

**Table 8. Predicted Fraction of Employees at Levels 1, 2, 3 and Leaving the Firm**

| **Years Since Entry** | **Level 1** ($a_\beta = 1$) ($b_\beta = 1$) | **Level 1** ($a_\beta = 50$) ($b_\beta = 50$) | **Level 2** ($a_\beta = 1$) ($b_\beta = 1$) | **Level 2** ($a_\beta = 50$) ($b_\beta = 50$) | **Level 3** ($a_\beta = 1$) ($b_\beta = 1$) | **Level 3** ($a_\beta = 50$) ($b_\beta = 50$) | **Exit** ($a_\beta = 1$) ($b_\beta = 1$) | **Exit** ($a_\beta = 50$) ($b_\beta = 50$) |
|---|---|---|---|---|---|---|---|---|
| 0 | 1.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0.330 | 0.625 | 0.262 | 0.291 | 0.000 | 0.000 | 0.408 | 0.085 |
| 2 | 0.058 | 0.080 | 0.087 | 0.093 | 0.168 | 0.186 | 0.686 | 0.641 |
| 3 | 0.071 | 0.103 | 0.021 | 0.032 | 0.115 | 0.119 | 0.793 | 0.746 |
| 4 | 0.010 | 0.000 | 0.020 | 0.029 | 0.064 | 0.740 | 0.906 | 0.898 |
| 5 | 0.009 | 0.013 | 0.003 | 0.000 | 0.039 | 0.048 | 0.949 | 0.939 |
| 6 | 0.004 | 0.006 | 0.003 | 0.004 | 0.019 | 0.021 | 0.974 | 0.969 |
| 7 | 0.001 | 0.001 | 0.001 | 0.002 | 0.010 | 0.012 | 0.988 | 0.985 |
| 8 | 0.001 | 0.001 | 0.000 | 0.000 | 0.005 | 0.006 | 0.994 | 0.992 |
| 9 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 | 0.003 | 0.997 | 0.997 |

### 6.4 Experiment 3: Increased Accuracy of Performance Monitoring

Suppose now that the probability of a high performance rating is one for a high ability worker and zero for a low ability worker, at either Level 1 or Level 2.[31] Since in the model the firm's production and monitoring technology coincide, a change in the probability of success for each type amounts to a change in the one-period expected revenue at either Levels 1 or 2, given that the same output realizations occur with different probabilities, as well as in the expected continuation profit from either level, given that the variance in posterior beliefs, when the output signal is perfectly informative, is maximal.

[31] In the simulation of the two experiments, the probabilities of success were set to $\alpha_1 = 9 \cdot 10^{-7}$ and $\beta_1 = 10^{-7}$, in case Level 1 is assumed to be perfectly informative about the worker's ability, and to $\alpha_2 = 9 \cdot 10^{-7}$ and $\beta_2 = 10^{-7}$, in case Level 2 is supposed to be perfectly informative.

**Table 9. Predicted Fraction of Employees at Levels 1, 2, 3 and Leaving the Firm**

| **Years (Since Entry)** | **Level 1** | **Level 1** ($\alpha_1 = 1$) ($\beta_1 = 0$) | **Level 2** | **Level 2** ($\alpha_1 = 1$) ($\beta_1 = 0$) | **Level 3** | **Level 3** ($\alpha_1 = 1$) ($\beta_1 = 0$) | **Exit** | **Exit** ($\alpha_1 = 1$) ($\beta_1 = 0$) |
|---|---|---|---|---|---|---|---|---|
| 0 | 1.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0.330 | 0.000 | 0.262 | 0.000 | 0.000 | 0.462 | 0.408 | 0.538 |
| 2 | 0.058 | 0.000 | 0.087 | 0.000 | 0.168 | 0.203 | 0.686 | 0.797 |
| 3 | 0.071 | 0.000 | 0.021 | 0.000 | 0.115 | 0.089 | 0.793 | 0.911 |
| 4 | 0.010 | 0.000 | 0.020 | 0.000 | 0.064 | 0.039 | 0.906 | 0.961 |
| 5 | 0.009 | 0.000 | 0.003 | 0.000 | 0.039 | 0.017 | 0.949 | 0.983 |
| 6 | 0.004 | 0.000 | 0.003 | 0.000 | 0.019 | 0.008 | 0.974 | 0.992 |
| 7 | 0.001 | 0.000 | 0.001 | 0.000 | 0.010 | 0.003 | 0.988 | 0.997 |
| 8 | 0.001 | 0.000 | 0.000 | 0.000 | 0.005 | 0.001 | 0.994 | 0.999 |
| 9 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 | 0.001 | 0.997 | 0.999 |

Compare the distribution of workers across levels when Level 1 is perfectly informative about the worker's true ability and in the benchmark case. As expected, the fact that observing a worker's performance for one period at Level 1 perfectly reveals his human capital makes the use of Level 1 profitable only the first year after entry. In case the worker is of low ability, in fact, the firm is better off by firing him than employing him at Level 1 afterwards, while, if the worker is of high ability, assigning him to Level 3 is for the firm the best alternative. As a consequence, then, the fraction of employees terminated is higher than in the benchmark case. Similarly, given that there is no informational value for the firm from assigning the worker to Level 2, after one period at Level 1 retained employees are only assigned to Level 3. This follows from the fact that $y_3(\overline{\theta}) = 40,842.562 > y_2(\overline{\theta}) = 1,472.475$, i.e., a high ability worker is more profitable for the firm when assigned to Level 3 than to Level 2. In this case, the change in the value of information can be as large as 563 percent, and it decreases from approximately $1,112$ percent to 0 percent.

As it can be seen from Table 10, the pattern of assignments to Level 1 when, instead, Level 2 perfectly reveals a worker's true skill, is analogous to the case in which Level 1 is perfectly informative. Because of the gain from assigning a high ability worker to Level 2 or from dismissing a low ability worker, no employee is retained at Level 1 after the first period. As a difference from the previous case, though, the fraction of employees at Level 2 does not decrease, a part from the third year after entry. This is due to the fact that, given that $\beta_2 = 0.000$ and $\alpha_2$ is relatively large, observing a low rating at high beliefs (at which Level 2 is still the best assignment) implies a small belief revision. The corresponding increase in the value of information to the firm ranges from 480.6 percent, when $\phi = 0.005$, to approximately 0 percent, when the firm knows with certainty the worker's ability.

**Table 10. Predicted Fraction of Employees at Levels 1, 2, 3 and Leaving the Firm**

| Years Since Entry | Level 1 | Level 1 ($\alpha_2 = 1$) ($\beta_2 = 0$) | Level 2 | Level 2 ($\alpha_2 = 1$) ($\beta_2 = 0$) | Level 3 | Level 3 ($\alpha_2 = 1$) ($\beta_2 = 0$) | Exit | Exit ($\alpha_2 = 1$) ($\beta_2 = 0$) |
|---|---|---|---|---|---|---|---|---|
| 0 | 1.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0.330 | 0.000 | 0.262 | 0.164 | 0.000 | 0.000 | 0.408 | 0.836 |
| 2 | 0.058 | 0.000 | 0.087 | 0.103 | 0.168 | 0.000 | 0.686 | 0.897 |
| 3 | 0.071 | 0.000 | 0.021 | 0.103 | 0.115 | 0.000 | 0.793 | 0.897 |
| 4 | 0.010 | 0.000 | 0.020 | 0.103 | 0.064 | 0.000 | 0.906 | 0.897 |
| 5 | 0.009 | 0.000 | 0.003 | 0.103 | 0.039 | 0.000 | 0.949 | 0.897 |
| 6 | 0.004 | 0.000 | 0.003 | 0.103 | 0.019 | 0.000 | 0.974 | 0.897 |
| 7 | 0.001 | 0.000 | 0.001 | 0.103 | 0.010 | 0.000 | 0.988 | 0.897 |
| 8 | 0.001 | 0.000 | 0.000 | 0.103 | 0.005 | 0.000 | 0.994 | 0.897 |
| 9 | 0.000 | 0.000 | 0.000 | 0.103 | 0.002 | 0.000 | 0.997 | 0.897 |

### 6.5 Experiment 4: A Two-Job Hierarchy

The last experiment performed is to assume that the firm's hierarchy only consists of Level 1, the entry level, and Level 3. Recall that Level 3 is the most profitable job position if the worker is truly of high ability, but it is also the one at which the firm incurs the greatest one-period profit loss if the worker's actual ability is low. The experiment is performed by setting $\underline{y}_2 = \overline{y}_2 = \alpha_2 = \beta_2 = 0$.

**Table 11. Predicted Fraction of Employees at Levels 1, 2, 3 and Leaving the Firm**

| Years Since Entry | Level 1 (Three Levels) | Level 1 (Two Levels) | Level 3 (Three Levels) | Level 3 (Two Levels) | Exit (Three Levels) | Exit (Two Levels) |
|---|---|---|---|---|---|---|
| 0 | 1.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0.330 | 0.344 | 0.000 | 0.424 | 0.408 | 0.232 |
| 2 | 0.058 | 0.069 | 0.168 | 0.269 | 0.686 | 0.662 |
| 3 | 0.071 | 0.002 | 0.115 | 0.130 | 0.793 | 0.868 |
| 4 | 0.010 | 0.001 | 0.064 | 0.058 | 0.906 | 0.941 |
| 5 | 0.009 | 0.000 | 0.039 | 0.025 | 0.949 | 0.975 |
| 6 | 0.004 | 0.000 | 0.019 | 0.011 | 0.974 | 0.989 |
| 7 | 0.001 | 0.000 | 0.010 | 0.005 | 0.988 | 0.995 |
| 8 | 0.001 | 0.000 | 0.005 | 0.002 | 0.994 | 0.998 |
| 9 | 0.000 | 0.000 | 0.002 | 0.001 | 0.997 | 0.999 |

In this case, workers whose assessed ability is sufficiently high, so that they are retained at the firm, are assigned to Level 3 rather than to Level 1. Since at Level 3 performance outcomes are imperfect signals of ability, workers who receive low ratings are more likely to be terminated than in the benchmark case, given that on average retained employees are assigned to Level 3 for lower belief values, i.e., when the impact of a low output signal on posterior beliefs is still significant. This is then the reason why in the second year since entry employees exit more often than in the benchmark case, even if the proportion of employees assigned to Level 1 does not change.

The fact that, for intermediate belief values, the firm can only assign the worker to Level 3, with the risk of a greater output destruction than at Level 2 if the worker is of low ability (i.e., $y_3(\overline{\theta}) - y_3(\overline{\theta}) > y_2(\overline{\theta}) - y_2(\overline{\theta})$) makes the firm less willing to employ a worker than in the benchmark case (employment starts at $\phi = 0.053$, compared to $\phi = 0.053$ in the benchmark case), with a corresponding reduction in the value of information of the order of 100 percent, when employment starts to be profitable, to almost 0 percent, when $\phi = 1$.

# 7 Related Literature

There are several related strands of literature. A number of papers, following Jovanovic [1979a, 1979b], have applied the one-armed Bandit framework to the study of turnover across firms.[32] Seminal paper on multi-tasking is Holmström and Milgrom [1991]. They investigate a multi-task principal-agent model in the presence of moral hazard and interpret job design as an instrument to control incentives, rather than a mechanism to generate information about an agent's unobserved ability. The implications of their model for the theory of job design is that, if measurement errors (as captured by the noise to signal ratio of performance per unit of time at a task) are correlated across tasks, grouping tasks with different performance characteristics in the same job is optimal. In this case grouping tasks allows the use of comparative performance evaluation, which in turn helps reducing the risk premium incurred by the principal in providing incentives. In our framework, on the contrary, different jobs may consist of tasks with dependent measurement errors, given that a worker's unobserved ability is correlated across tasks and, being tasks dynamically complementary in information production, workers can be assigned to different tasks only as their tenure at the firm increases.

Among the contributions which analyze job assignment inside firms, Prendergast [1993] rationalizes promotions as an equilibrium device to reward the nonverifiable acquisition of firm-specific human capital. Waldman [1984b] focuses instead on the distortions in the equilibrium assignment process which arise when promotions serve as a public signal to the market about a worker's unobserved ability. Fairburn and Malcomson [2001] offer an interpretation for the relationship between performance, incentives and promotions based on the conflict of interest between managers and firm. Specifically, they study the relative incentive power of promotions and monetary transfers, when

[32] See also the discussion contained in the companion paper.

supervisors of employees are subject to influence activities on the part of employees, and show that the use of promotions reduces the incentive for managers to be affected by them.[33]

The paper closest in spirit to ours is Gibbons and Waldman [1999b]. They develop a model of learning, job assignment and human capital acquisition which accounts for a broad pattern of evidence on wage and promotion dynamics inside firms. They assume that there exists an output interaction between learning and human capital acquisition, which both affect a worker's expected product in a period. In their model, as in ours, an equilibrium hierarchy of job positions results from the assumption that higher ability is more valuable at higher level jobs. Since human capital is accumulated by experience, all workers eventually reach the highest job position in the firm's hierarchy as they age. Because of learning on the part of the firm and the accumulation of skills on the part of the worker, demotions are rare. The main differences between our framework and theirs are that in our case ($i$) the job performed by a worker affects the rate of learning about ability, and ($ii$) a worker of low ability is nonprofitable for the firm. Experimentation on ability affects then dynamic screening both through retention and job assignment. In particular, in our framework workers move up the job ladder purely as a consequence of the firm's improved estimate of their ability. However, because of the informational value of lower level jobs only when uncertainty about ability is highest, demotions can be rare.

On the empirical side, due to the confidentiality of the data required, only a few studies analyze intra-firm job transitions or wage dynamics. Baker, Gibbs and Holmström [1994a] provide a detailed case study analysis of the data from which our estimation sample has been selected, finding evidence for the hypothesis that a firm's internal hierarchy acts as an information acquisition filter, to screen employees according to their unobserved abilities. Baker, Gibbs and Holmström [1994b] test whether existing explanations for wage dynamics, specifically on-the-job training, learning and stationary incentive models, are consistent with the wage policy they infer from the data. They find that none of these models can alone be reconciled with the patterns emerging from the data. Chiappori, Salanié and Valentin [1999] consider a model of wage formation, in the presence of learning and downward wage rigidity, and find evidence of a 'late beginner' property in the dynamics of wages, i.e., after controlling for the wage at $t$, the wage at $t+1$ is negatively correlated with the wage at $t-1$. Finally, focussing on the analysis of turnover, Nagypal [2002] adopts a structural estimation approach to test the relative explanatory power of learning about ability versus learning on the job in determining the intertemporal profile of the hazard rate of employment termination, using a French matched employer-employee dataset. Her estimation results support a learning interpretation for inter-firm job transitions.[34]

[33] There is finally a number of papers which study the interaction of strategic aspects of the oligopoly problem with a decision maker's incentive to experiment and analyze the degree of efficiency of market experimentation. An example is Bergemann and Välimäki [1996].

[34] Kwon [2004], on the other hand, tries to assess the relative importance of sorting and incentive provision in shaping the dynamic profile of the probability of dismissal. His estimation results provide evidence for the incentive model.

# 8 Conclusion

This paper has developed a learning model of retention and job assignment and provided a structural estimation of it, using ten years of observations on level assignments and performance ratings for the cohorts of managers employed at a single U.S. service firm between 1970 and 1979. Estimation results confirm that a firm's internal hierarchy can act as an information acquisition filter, with performance at lower level jobs being used by the firm to learn about workers' true productivity, for the benefit of future assignment decisions. The sequential screening mechanism, which characterizes the firm's employment policy in the equilibria of interest, has been shown to be also a property of the promotion dynamics estimated from the data. In particular, the estimated retention and task assignment policy is the one predicted by the model. Estimation results confirm that the model succeeds in fitting the probability of retention and promotion at the job positions at which most managers are employed over the sample period.

A number of stylized facts have been documented in the literature on the internal economics of the firm about the dynamics of wages and promotions (see Gibbons and Waldman [1999a, 1999b] for a comprehensive reference). The interaction of outside labor market competition with a firm's incentive to experiment on workers' ability is an important determinant of job dynamics in firms, given its impact on wages and, therefore, on the profitability of employment at any job.[35] An issue of interest within a learning framework is also the extent to which the gradual assignment of employees to higher level jobs, at which ability is more valuable, is the result of firm's learning about workers' ability or can be attributed to workers acquiring new skills on the job. The exploration of these issues constitutes the specific object of present and future research.

# Appendix A

**Proof of Proposition 1**: The fact that $\Pi(\cdot)$ is well-defined and continuous can be shown by a standard Contraction Mapping argument. Under (A1) and (A2), it can also be shown that it is increasing. As for the convexity of $\Pi(\cdot)$, the proof is adapted from Banks and Sundaram [1992a]. Recall

$$\begin{aligned}\Pi(\phi) \;=\; & \max\{\overline{\Pi}, p_1(\phi)[\overline{y}_1 + \delta(1-\xi_1)\Pi(\phi_{1h}(\phi))] + (1-p_1(\phi))[\underline{y}_1 + \delta(1-\xi_1)\Pi(\phi_{1l}(\phi))] + \delta\xi_1\overline{\Pi},\\ & p_2(\phi)[\overline{y}_2 + \delta(1-\xi_2)\Pi(\phi_{2h}(\phi))] + (1-p_2(\phi))[\underline{y}_2 + \delta(1-\xi_2)\Pi(\phi_{2l}(\phi))] + \delta\xi_2\overline{\Pi},\\ & p_3(\phi)[\overline{y}_3 + \delta(1-\xi_3)\Pi(\phi_{3h}(\phi))] + (1-p_3(\phi))[\underline{y}_3 + \delta(1-\xi_3)\Pi(\phi_{3l}(\phi))] + \delta\xi_3\overline{\Pi}\}.\end{aligned}$$

[35] See the comment in Subsection 4.2 on the current estimation of the wage process.

Define the mappings $T_k$, $k = 1, 2, 3$, and $T$ on $\mathcal{C}[0, 1]$, the space of continuous functions on the unit interval, as follows. For $k = 1, 2, 3$, let

$$\begin{aligned} T_k f(\phi) &= p_k(\phi)[\overline{y}_k + \delta(1-\xi_k) f(\phi_{kh}(\phi))] + (1 - p_k(\phi))[\underline{y}_k + \delta(1-\xi_k) f(\phi_{kl}(\phi))] + \delta\xi_k\overline{\Pi} \\ &= p_k(\phi)\overline{y}_k + (1-p_k(\phi))\underline{y}_k + \delta G_k f(\phi), \\ G_k f(\phi) &= (1-\xi_k)[p_k(\phi) f(\phi_{kh}(\phi)) + (1-p_k(\phi)) f(\phi_{kl}(\phi))] + \xi_k\overline{\Pi} \end{aligned}$$

and $Tf(\phi) = \max\{\overline{\Pi}, T_1 f(\phi), T_2 f(\phi), T_3 f(\phi)\}$. We will proceed in two steps. We will first show that, if $f$ is convex, then $Tf$ is also convex. Let $\phi', \phi'' \in [0, 1]$, $\lambda \in (0, 1)$ and $\phi^* \equiv (1-\lambda)\phi' + \lambda\phi''$. Define $e(y_{kh}) \equiv \frac{\lambda p_k(\phi'')}{p_k(\phi^*)} \in (0, 1)$ and $e(y_{kl}) \equiv \frac{\lambda(1-p_k(\phi''))}{1-p_k(\phi^*)} \in (0, 1)$. Equivalently, $1 - e(y_{kh}) \equiv \frac{(1-\lambda)p_k(\phi')}{p_k(\phi^*)} \in (0, 1)$ and $1 - e(y_{kl}) \equiv \frac{(1-\lambda)(1-p_k(\phi'))}{1-p_k(\phi^*)} \in (0, 1)$. Note that

$$\begin{aligned} (1 - e(y_{kh}))\phi_{kh}(\phi') + e(y_{kh})\phi_{kh}(\phi'') &= \frac{(1-\lambda)\alpha_k\phi' + \lambda\alpha_k\phi''}{p_k(\phi^*)} = \phi_{kh}(\phi^*) \\ (1 - e(y_{kl}))\phi_{kl}(\phi') + e(y_{kl})\phi_{kl}(\phi'') &= \frac{(1-\lambda)(1-\alpha_k)\phi' + \lambda(1-\alpha_k)\phi''}{1-p_k(\phi^*)} = \phi_{kl}(\phi^*) \end{aligned}$$

by definition of $\phi_{kh}(\phi)$ and $\phi_{kl}(\phi)$, $k = 1, 2, 3$. Suppose $f$ is convex. For $k = 1, 2, 3$, from Jensen's inequality for convex functions, it follows

$$\begin{aligned} G_k f(\phi^*) &= (1-\xi_k)[p_k(\phi^*) f(\phi_{kh}(\phi^*)) + (1-p_k(\phi^*)) f(\phi_{kl}(\phi^*))] + \xi_k\overline{\Pi} \\ &\leq (1-\xi_k)p_k(\phi^*)[(1-e(y_{kh})) f(\phi_{kh}(\phi')) + e(y_{kh}) f(\phi_{kh}(\phi''))] \\ &\quad + (1-\xi_k)(1-p_k(\phi^*))[(1-e(y_{kl})) f(\phi_{kl}(\phi')) + e(y_{kl}) f(\phi_{kl}(\phi''))] + \xi_k\overline{\Pi}. \end{aligned}$$

Note that $p_k(\phi^*)(1 - e(y_{kh})) = (1-\lambda)p_k(\phi')$, $p_k(\phi^*)e(y_{kh}) = \lambda p_k(\phi'')$, $(1 - p_k(\phi^*))(1 - e(y_{kl})) = (1-\lambda)(1-p_k(\phi'))$ and $(1 - p_k(\phi^*))e(y_{kl}) = \lambda(1 - p_k(\phi''))$. Rearranging terms,

$$\begin{aligned} G_k f(\phi^*) &\leq (1-\xi_k)[(1-\lambda)p_k(\phi') f(\phi_{kh}(\phi')) + \lambda p_k(\phi'') f(\phi_{kh}(\phi''))] \\ &\quad + (1-\xi_k)[(1-\lambda)(1-p_k(\phi')) f(\phi_{kl}(\phi')) + \lambda(1-p_k(\phi'')) f(\phi_{kl}(\phi''))] + \xi_k\overline{\Pi} \\ &= (1-\lambda)[(1-\xi_k)p_k(\phi') f(\phi_{kh}(\phi')) + (1-\xi_k)(1-p_k(\phi')) f(\phi_{kl}(\phi')) + \xi_k\overline{\Pi}] \\ &\quad + \lambda[(1-\xi_k)p_k(\phi'') f(\phi_{kh}(\phi'')) + (1-\xi_k)(1-p_k(\phi'')) f(\phi_{kl}(\phi'')) + \xi_k\overline{\Pi}] \\ &= (1-\lambda)G_k f(\phi') + \lambda G_k f(\phi''). \end{aligned}$$

Observe that $p_k(\phi^*) = (1-\lambda)p_k(\phi') + \lambda p_k(\phi'')$ and $1 - p_k(\phi^*) = (1-\lambda)(1-p_k(\phi')) + \lambda(1-p_k(\phi''))$. From this,

$$\begin{aligned} T_k f(\phi^*) &= p_k(\phi^*)\overline{y}_k + (1-p_k(\phi^*))\underline{y}_k + \delta G_k f(\phi^*) \\ &\leq [(1-\lambda)p_k(\phi') + \lambda p_k(\phi'')]\overline{y}_k + [(1-\lambda)(1-p_k(\phi')) + \lambda(1-p_k(\phi''))]\underline{y}_k \\ &\quad + \delta(1-\lambda)G_k f(\phi') + \delta\lambda G_k f(\phi'') \\ &= (1-\lambda)T_k f(\phi') + \lambda T_k f(\phi''). \end{aligned}$$

As the maximum of convex functions, $Tf$ is convex whenever $f$ is convex.

As for the second step, we will prove that the (unique) fixed point of the mapping $T$ is also convex. Let $\mathcal{CX}$ be the set of all convex functions $f$ such that $f \leq Tf$. Note that $\mathcal{CX}$ is bounded above and non-empty. Let $f^* \equiv \sup\{f(\phi)|f \in \mathcal{CX}\}$. As the pointwise supremum of convex functions, $f^*$ is convex. Observe that $T$ is a monotone increasing operator. Then, $f^*(\phi) = \sup\{f(\phi)|f \in \mathcal{CX}\} \leq \sup\{Tf(\phi)|f \in \mathcal{CX}\}$, by definition of $\mathcal{CX}$. Also, by monotonicity of $T$, $\sup\{Tf(\phi)|f \in \mathcal{CX}\} \leq Tf^*(\phi)$. These combined observations imply $f^* \leq Tf^*$ or $f^* \in \mathcal{CX}$. Recall that, for all $f \in \mathcal{CX}$, $f \leq Tf$. Thus, by monotonicity of $T$, $Tf \leq T(Tf)$, which implies $Tf \in \mathcal{CX}$ if $f \in \mathcal{CX}$. In particular, $Tf^* \in \mathcal{CX}$. Therefore, by the definition of $f^*$, it must be $f^* \geq Tf^*$. This, together with $f^* \leq Tf^*$, yields $Tf^* = f^*$ or, equivalently, $f^*$ is a fixed point of the mapping $T$. But since $T$ is a contraction, it has a unique fixed point. This completes the proof of the claim.

**Proof of Proposition 2**: Let $\phi_1^*$ be the cut-off belief value which makes the firm indifferent between not hiring the worker and employing him at task 1, i.e., $\overline{\Pi} = y_1(\phi_1^*) - U + \delta E_1 \Pi(\phi_1^*)$. Notice that $\phi'' > \phi'$ implies that the distribution of the updated posterior conditional on $\phi''$, after revenue realizes at a job, first-order stochastically dominates the one conditional on $\phi'$. Then, $E_1\Pi(\cdot)$ is increasing in $\phi$ if $\Pi(\cdot)$ is increasing. Since $y_1(\cdot)$ is strictly increasing, it follows that $\Pi_1(\cdot)$ is also strictly increasing (by a similar argument, it can be shown that $\Pi_2(\cdot)$ and $\Pi_3(\cdot)$ are strictly increasing in $\phi$ as well). Thus, $\phi_1^*$ is uniquely determined and, with $y_1(\overline{\theta}) > \Pi > y_1(\underline{\theta})$, $\phi_1^* \in (0,1)$. Suppose now that the following condition holds

$$\Pi_1(\phi_1^*) = y_1(\phi_1^*) - U + \delta E_1 \Pi(\phi_1^*) > \Pi_2(\phi_1^*) = y_2(\phi_1^*) - U + \delta E_2 \Pi(\phi_1^*). \tag{4}$$

Then, together with $y_2(\overline{\theta}) > \overline{\Pi} > y_2(\underline{\theta})$, condition (4) yields

$$\Pi_2(\overline{\theta}) \equiv y_2(\overline{\theta}) - U + \frac{y_3(\overline{\theta}) - U + \delta\xi_3\overline{\Pi}}{1-\delta(1-\xi_3)} > \overline{\Pi} > \Pi_2(\phi_1^*) = y_2(\phi_1^*) - U + \delta E_2 \Pi(\phi_1^*)$$

which implies that there exists a unique value $\phi_{0,2}^* \in (0,1)$, with $\phi_{0,2}^* > \phi_1^*$, such that $\overline{\Pi} = \Pi_2(\phi_{0,2}^*)$. Moreover, since $y_2(\overline{\theta}) > y_1(\overline{\theta})$, from (4) and

$$\Pi_2(\overline{\theta}) > \Pi_1(\overline{\theta}) = y_1(\overline{\theta}) - U + \frac{\delta[y_3(\overline{\theta}) - U + \delta\xi_3\overline{\Pi}]}{1-\delta(1-\xi_3)}$$

it follows that there exists a value $\phi_2^* \in (0,1)$, with $\phi_2^* > \phi_1^*$, satisfying

$$y_1(\phi_2^*) - U + \delta E_1 \Pi(\phi_2^*) = y_2(\phi_2^*) - U + \delta E_2 \Pi(\phi_2^*)$$

so that tasks 1 and 2 are equally profitable. Since $\phi_2^* > \phi_1^*$, by definition of $\phi_1^*$ it follows $\Pi(\phi_2^*) > \overline{\Pi}$ and, then, $\phi_2^* > \phi_{0,2}^*$. Suppose now that the following condition holds as well

$$\Pi_2(\phi_2^*) = y_2(\phi_2^*) - U + \delta E_2 \Pi(\phi_2^*) > \Pi_3(\phi_2^*) = y_3(\phi_2^*) - U + \delta E_3 \Pi(\phi_2^*). \tag{5}$$

With $y_3(\overline{\theta}) > y_2(\overline{\theta})$ and $\phi_2^* < 1$,

$$\Pi_3(\overline{\theta}) \equiv \frac{y_3(\overline{\theta}) - U + \delta\xi_3\overline{\Pi}}{1-\delta(1-\xi_3)} > \Pi_2(\phi_2^*) = y_2(\phi_2^*) - U + \delta E_2 \Pi(\phi_2^*)$$

which, together with (5), implies that there exists a value $\phi_3^* \in (0,1)$, with $\phi_3^* > \phi_2^*$, satisfying

$$y_2(\phi_3^*) - U + \delta E_2\Pi(\phi_3^*) = y_3(\phi_3^*) - U + \delta E_3\Pi(\phi_3^*) > \overline{\Pi}.$$

Observe that, if the difference $\Pi_2(\phi) - \Pi_1(\phi)$ is strictly increasing, when $\phi \in [\phi_1^*, \phi_3^*]$, and the difference $\Pi_3(\phi) - \Pi_2(\phi)$ is strictly increasing, when $\phi \in [\phi_2^*, 1]$, then the cut-off values $\phi_2^*$ and $\phi_3^*$ are uniquely determined. What we will show next is first that, under the conditions stated in the Proposition, (4) and (5) hold and, then, that $\Pi_2(\phi)-\Pi_1(\phi)$ and $\Pi_3(\phi)-\Pi_2(\phi)$ are strictly increasing in $\phi$ over the specified belief ranges.

Let $k = 1, 2$. Notice that, if $\alpha_k\beta_{k+1} > \alpha_{k+1}\beta_k$, $\phi_{hk}(\phi) > \phi_{hk+1}(\phi)$ for $\phi \in (0,1)$. Similarly, $\alpha_{k+1} > \beta_{k+1}$ and $\alpha_k > \beta_k$ imply, respectively, $\phi_{hk+1}(\phi) > \phi_{lk+1}(\phi)$ and $\phi_{hk}(\phi) > \phi_{lk}(\phi)$, if $\phi \in (0,1)$. Moreover, if $(1-\alpha_{k+1})(1-\beta_k) > (1-\alpha_k)(1-\beta_{k+1})$, it follows $\phi_{lk+1}(\phi) > \phi_{lk}(\phi)$ for $\phi \in (0,1)$. A sufficient condition for $\alpha_k\beta_{k+1} > \alpha_{k+1}\beta_k$ and $(1-\alpha_{k+1})(1-\beta_k) > (1-\alpha_k)(1-\beta_{k+1})$ to hold is $\alpha_k \geq \alpha_{k+1}$ and $\beta_{k+1} \geq \beta_k$. Consider now the distributions of the next period value of $\phi$, $\phi'$, conditional on its current period value and the worker being assigned to tasks $k$ or $k+1$. Denote the two corresponding cumulative distribution functions, respectively, by $F(\phi';k)$ and $G(\phi';k+1)$. Observe that the mean of the two distributions is $\phi$. Now, the fact that $\phi_{hk}(\phi) > \phi_{hk+1}(\phi) > \phi_{lk+1}(\phi) > \phi_{lk}(\phi)$, and $F(\phi';k)$ and $G(\phi';k+1)$ are two-outcome distributions, implies that $F(\phi';k)$ constitutes a mean-preserving spread of $G(\phi';k+1)$. Equivalently, $G(\phi';k+1)$ second-order stochastically dominates $F(\phi';k)$. By definition, for any two distributions $F(x)$ and $G(x)$ with the same mean, $G$ second-order stochastically dominates $F$ if $\int \psi(x)dF(x) \geq \int \psi(x)dG(x)$ for every increasing convex function $\psi : \mathbb{R}_+ \to \mathbb{R}$. It then follows $E_k\Pi(\phi) \geq E_{k+1}\Pi(\phi)$, by convexity of $\Pi(\cdot)$, if the exogenous separation rates $\xi_k$ and $\xi_{k+1}$ are sufficiently small. This argument, for $k = 1, 2$, ensures that $E_1\Pi(\phi) \geq E_2\Pi(\phi) \geq E_3\Pi(\phi)$. Observe now that $E_1\Pi \geq \overline{\Pi}$ implies $\phi_1^* \leq \phi_{0,1}$. The condition $\phi_{0,1} < \phi_{1,2}$ in turn implies $y_1(\phi_{0,1}) > y_2(\phi_{0,1})$. Note that $y_2(\overline{\theta}) > y_1(\overline{\theta})$ and $y_1(\underline{\theta}) > y_2(\underline{\theta})$ imply that the difference $y_1(\phi) - y_2(\phi)$ is strictly decreasing. With $\phi_1^* < \phi_{0,1}$, from $y_1(\phi_{0,1}) > y_2(\phi_{0,1})$ it follows $y_1(\phi_1^*) > y_2(\phi_1^*)$. Then, condition (4) holds true. Recall that $\overline{\phi}$ is the belief value for which the firm is indifferent, in the static case, between tasks 1 and 2, if task 1 is perfectly informative about ability, while task 2 is completely uninformative. Then, $\overline{\phi}$ can be computed as

$$\overline{\phi} \equiv \frac{y_1(\underline{\theta}) - y_2(\underline{\theta}) + \frac{\delta(1-\xi_2)\Pi}{1-\delta(1-\xi_2)}}{\frac{y_2(\overline{\theta})-y_2(\underline{\theta})}{1-\delta(1-\xi_2)} - y_1(\overline{\theta}) + y_1(\underline{\theta}) - \frac{\delta(1-\xi_1)(y_3(\overline{\theta})-U-\Pi)}{1-\delta(1-\xi_3)}}.$$

Note that $\overline{\phi} \in (0,1)$, if $\xi_k$, $k = 1, 2, 3$, is sufficiently small. Also, $\phi_2^* < \overline{\phi}$. For $y_3(\overline{\theta}) > y_2(\overline{\theta})$ and $y_2(\underline{\theta}) > y_3(\underline{\theta})$, the difference $y_2(\phi) - y_3(\phi)$ is strictly decreasing. Moreover, $\phi_2^* < \overline{\phi}$ implies that, if $y_2(\overline{\phi}) - y_3(\overline{\phi}) > 0$, then $y_2(\phi_2^*) - y_3(\phi_2^*) > 0$. Observe finally that $y_2(\overline{\phi}) - y_3(\overline{\phi}) > 0$ is equivalent to $y_2(\underline{\theta}) - y_3(\underline{\theta}) > k(\overline{\phi})[y_3(\overline{\theta}) - y_2(\overline{\theta})]$. The fact that $\phi_1^* < \phi_{0,1}$, $\phi_2^* > \phi_{1,2}$ and $\phi_3^* > \phi_{2,3}$ is consequence that at $\phi_1^*$, $\phi_2^*$ and $\phi_3^*$ the firm's value function is kinked, so that, respectively, $E_1\Pi(\cdot) > \overline{\Pi}$, $E_1\Pi(\cdot) > E_2\Pi(\cdot)$ and $E_2\Pi(\cdot) > E_3\Pi(\cdot)$.

We will now show that the cut-off belief values $\phi_2^*$ and $\phi_3^*$ are uniquely determined. Define, analogously to the proof of Proposition 1, $Tf(\phi) = \max\{\overline{\Pi}, T_1f(\phi), T_2f(\phi), T_3f(\phi)\}$, where, for

$k = 1, 2, 3,$

$$T_k f(\phi) = p_k(\phi)[\overline{y}_k + \delta(1-\xi_k) f(\phi_{kh}(\phi))] + (1 - p_k(\phi))[\underline{y}_k + \delta(1-\xi_k) f(\phi_{kl}(\phi))] + \delta\xi_k \overline{\Pi}.$$

Suppose that $T_2 f - T_1 f$ and $T_3 f - T_2 f$ are increasing in $\phi$. From $E_1 f \geq E_2 f \geq E_3 f$, and $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$, it follows that, when $T_1 f < T_0 f$, also $T_2 f < T_0 f$ and $T_3 f < T_0 f$. The fact that $y_1(\overline{\theta}) > \Pi > y_1(\underline{\theta})$ implies that there exists a $\hat{\phi}_1 \in (0, 1)$ such that $T_1 f = \overline{\Pi}$. Since $E_1 f \geq E_2 f \geq E_3 f$, and $\phi_{0,1} < \phi_{1,2} < \phi_{2,3}$, it follows that, at $\hat{\phi}_1$, $T_1 f \geq T_2 f$ implies $T_1 f \geq T_3 f$. Therefore, $Tf(\hat{\phi}_1) = T_1 f(\hat{\phi}_1)$. From the fact that $T_2 f - T_1 f$ is increasing, and that $y_2(\overline{\theta}) > y_1(\overline{\theta})$, it also follows $Tf = T_1 f$ for $\phi \in [\hat{\phi}_1, \hat{\phi}_2)$, for some $\hat{\phi}_2 < 1$. Then, for $\phi \geq \hat{\phi}_2$, $Tf = T_2 f$ or $Tf = T_3 f$. Now, $y_2(\overline{\phi}) > y_3(\overline{\phi})$, with $\overline{\phi} > \hat{\phi}_2$, and $E_2 f \geq E_3 f$ yield that, at $\hat{\phi}_2$, $Tf = T_2 f$. Since $T_3 f - T_2 f$ is increasing and $y_3(\overline{\theta}) > y_2(\overline{\theta})$, it follows that there also exists $\hat{\phi}_3$ such that $Tf = T_2 f$ for $\phi \in [\hat{\phi}_2, \hat{\phi}_3)$ and $Tf = T_3 f$ for $\phi \geq \hat{\phi}_3$. Then, $Tf$ can be rewritten as

$$Tf = \begin{cases} T_3 f, & \text{if } \phi \in [\hat{\phi}_3, 1] \\ T_2 f, & \text{if } \phi \in [\hat{\phi}_2, \hat{\phi}_3) \\ T_1 f, & \text{if } \phi \in [\hat{\phi}_1, \hat{\phi}_2) \\ T_0 f, & \text{if } \phi \in [0, \hat{\phi}_1). \end{cases}$$

Consider now the differences $T_3(Tf) - T_2(Tf)$ and $T_2(Tf) - T_1(Tf)$. Notice that they can be rewritten, respectively, as[36]

$$T_3(Tf) - T_2(Tf) = [T_3(Tf) - T_3(T_2 f)] + [T_2(T_3 f) - T_2(Tf)] + [T_3(T_2 f) - T_2(T_3 f)] \tag{6}$$

$$T_2(Tf) - T_1(Tf) = [T_2(Tf) - T_2(T_1 f)] + [T_1(T_2 f) - T_1(Tf)] + [T_2(T_1 f) - T_1(T_2 f)]. \tag{7}$$

Suppose that, for any real-valued function $f$ on $[0, 1]$, $T_3 f - T_2 f$ increasing over $[\hat{\phi}_2, 1]$ implies that $T_3(Tf) - T_2(Tf)$ is strictly increasing for the same values of $\phi$ and, similarly, that $T_2 f - T_1 f$ increasing over $[\hat{\phi}_1, \hat{\phi}_3]$ implies that $T_2(Tf) - T_1(Tf)$ is strictly increasing for the same values of $\phi$. Then, since $\Pi(\cdot)$ is the unique fixed point of $T$, $\Pi_3(\phi) - \Pi_2(\phi)$ and $\Pi_2(\phi) - \Pi_1(\phi)$ must be strictly increasing over those belief ranges. To prove that $T_3 f - T_2 f$ increasing implies that $T_3(Tf) - T_2(Tf)$ is strictly increasing and, similarly, that $T_2 f - T_1 f$ increasing implies that $T_2(Tf) - T_1(Tf)$ is strictly increasing, it is enough to show that each term in the right-hand side of (6) and (7) is increasing, and at least one strictly increasing, over the desired belief range. Notice that (6) can then be rewritten as

$$\begin{aligned} 1_{\{\phi \in [\hat{\phi}_2, 1]\}} T_3(Tf) \; - \; T_2(Tf) &= 1_{\{\phi \in [\hat{\phi}_3, 1]\}} \{T_3(T_3 f) - T_3(T_2 f)\} + 1_{\{\phi \in [\hat{\phi}_2, \hat{\phi}_3]\}} \{T_2(T_3 f) - T_2(T_2 f)\} \\ &\quad + 1_{\{\phi \in [\hat{\phi}_2, 1]\}} (1-\delta)^2 (y_3(\phi) - y_2(\phi)) \\ &= 1_{\{\phi \in [\hat{\phi}_3, 1]\}} \{\delta E_3 [T_3 f - T_2 f]\} + 1_{\{\phi \in [\hat{\phi}_2, \hat{\phi}_3]\}} \{\delta E_2 [T_3 f - T_2 f]\} \\ &\quad + 1_{\{\phi \in [\hat{\phi}_2, 1]\}} (1-\delta)^2 (y_3(\phi) - y_2(\phi)). \end{aligned} \tag{8}$$

The first equality follows from the definition of $Tf$ and the exchangeability of the output signal, which implies $E_3 E_2 f = E_2 E_3 f$. With $T_3 f - T_2 f$ increasing, and the fact that $\phi'' > \phi'$ implies that

[36] The decomposition is analogous to the one used in the proof of Lemma 1 in Kakigi [1983].

the distribution of the updated posterior at any job, conditional on $\phi''$, first-order stochastically dominates the one conditional on $\phi'$, it follows that the first two terms in (8) are increasing. Since the difference $y_3(\phi) - y_2(\phi)$ is strictly increasing, it follows $T_3(Tf) - T_2(T_2 f)$ is strictly increasing as well for $\phi \in [\phi_2^*, 1]$. Similarly, when $\phi \in [\hat{\phi}_1, \hat{\phi}_3]$, (7) can be rewritten as

$$\begin{aligned}
1_{\{\phi\in[\hat{\phi}_1,\hat{\phi}_3]\}} T_2(Tf) \; - \; T_1(Tf) &= 1_{\{\phi\in[\hat{\phi}_2,\hat{\phi}_3]\}}\{T_2(T_2 f) - T_2(T_1 f)\} + 1_{\{\phi\in[\hat{\phi}_1,\hat{\phi}_2)\}}\{T_1(T_2 f) - T_1(T_1 f)\} \\
&\quad + 1_{\{\phi\in[\hat{\phi}_1,\hat{\phi}_3]\}}(1-\delta)^2 (y_2(\phi) - y_1(\phi)) \\
&= 1_{\{\phi\in[\hat{\phi}_2,\hat{\phi}_3]\}}\{\delta E_2[T_2 f - T_1 f] + 1_{\{\phi\in[\hat{\phi}_1,\hat{\phi}_2)\}}\delta E_1[T_2 f - T_1 f] \\
&\quad + 1_{\{\phi\in[\hat{\phi}_1,\hat{\phi}_3]\}}(1-\delta)^2 (y_2(\phi) - y_1(\phi)).
\end{aligned}$$

As before, the first equality follows from the definition of $Tf$ and the fact that $E_2 E_1 f = E_1 E_2 f$. With $T_2 f - T_1 f$ increasing, it follows that $E_k[T_2 f - T_1 f]$ is also increasing, for $k = 1, 2$. Since $y_2(\phi) - y_1(\phi)$ is strictly increasing, the difference $T_2(Tf) - T_1(T_2 f)$ is strictly increasing for $\phi \in [\phi_1^*, \phi_3^*]$. This completes the proof of the claim.

As discussed in the companion paper, the game between the firm and the worker admits a representation as a complete information game, with a unique starting node given by the the firm and worker' prior, and a perfectly observed move by Nature in each period, which determines the known transition on the state variable, $\phi$. An alternative representation is an incomplete information game where Nature moves first selecting the type of the worker. For the characterization of the equilibrium outcomes of interest, the two representations are equivalent. Briefly, the result hinges upon the equivalence between Perfect Bayes Nash equilibria and MPE's, which derives in this framework from the fact that the worker's acceptance behavior in a period does not affect his expected discounted lifetime income. □

**Proof of Proposition 3**: The argument can be adapted from the proof of Proposition 6 in the companion paper. □

# Appendix B

## B.1 Sample Without Level or Performance Rating Missing

**Table B1. Fraction of High Ratings (Original Sample (OS), Estimation Sample (ES))**

| Rating | OS (Total) | OS | ES (Total) | ES |
|---|---|---|---|---|
| 1 | 13,994 | 0.306 | 763 | 0.398 |
| 2 | 22,756 | 0.498 | 954 | 0.497 |
| 3 | 8,455 | 0.185 | 197 | 0.103 |
| 4 | 438 | 0.010 | 2 | 0.001 |
| 5 | 30 | 0.001 | 3 | 0.002 |
| Total | 45,673 | 1.000 | 1,919 | 1.000 |

**Table B2. Distribution of Employees Across Levels (12 Years of Education at Entry - 43 Employees)**

| Years (Since Entry) | Level 1 | Level 2 | Level 3 | Exit | Total |
|---|---|---|---|---|---|
| 0 | 1.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| 1 | 0.698 | 0.116 | 0.000 | 0.186 | 1.000 |
| 2 | 0.465 | 0.093 | 0.070 | 0.372 | 1.000 |
| 3 | 0.233 | 0.140 | 0.093 | 0.535 | 1.000 |
| 4 | 0.047 | 0.140 | 0.116 | 0.698 | 1.000 |
| 5 | 0.000 | 0.070 | 0.047 | 0.884 | 1.000 |
| 6 | 0.000 | 0.023 | 0.000 | 0.977 | 1.000 |

**Table B3. Hazard Rates of Exit and Promotion by Level (12 Years of Education at Entry - 43 Employees)**

| Years (at Level) | Level 1 to Exit | Level 1 to Level 2 | Level 2 to Exit | Level 2 to Level 3 | Level 3 to Exit |
|---|---|---|---|---|---|
| 1 | 0.186 | 0.116 | 0.200 | 0.400 | 0.000 |
| 2 | 0.233 | 0.067 | 0.000 | 0.500 | 0.000 |
| 3 | 0.350 | 0.150 | 0.000 | 0.000 | 0.667 |
| 4 | 0.500 | 0.300 | 0.000 | 0.000 | 1.000 |
| 5 | 1.000 | 0.000 | 1.000 | 0.000 | - |
| 6 | - | - | - | - | - |

**Table B4. Distribution of Employees Across Levels (13-15 Years of Education at Entry - 70 Employees)**

| Years (Since Entry) | Level 1 | Level 2 | Level 3 | Exit | Total |
|---|---|---|---|---|---|
| 0 | 1.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| 1 | 0.700 | 0.100 | 0.000 | 0.200 | 1.000 |
| 2 | 0.300 | 0.143 | 0.014 | 0.543 | 1.000 |
| 3 | 0.171 | 0.100 | 0.043 | 0.686 | 1.000 |
| 4 | 0.114 | 0.057 | 0.071 | 0.757 | 1.000 |
| 5 | 0.043 | 0.057 | 0.014 | 0.886 | 1.000 |
| 6 | 0.014 | 0.043 | 0.014 | 0.929 | 1.000 |
| 7 | 0.014 | 0.029 | 0.029 | 0.929 | 1.000 |
| 8 | 0.000 | 0.000 | 0.014 | 0.986 | 1.000 |

**Table B5. Hazard Rates of Exit and Promotion by Level (13-15 Years of Education at Entry - 70 Employees)**

| Years (at Level) | Level 1 to Exit | Level 1 to Level 2 | Level 2 to Exit | Level 2 to Level 3 | Level 3 to Exit |
|---|---|---|---|---|---|
| 1 | 0.200 | 0.100 | 0.571 | 0.000 | 0.000 |
| 2 | 0.408 | 0.143 | 0.000 | 0.667 | 0.000 |
| 3 | 0.333 | 0.095 | 0.000 | 0.000 | 0.000 |
| 4 | 0.250 | 0.083 | 0.000 | 0.000 | 0.000 |
| 5 | 0.500 | 0.125 | 1.000 | 0.000 | 0.000 |
| 6 | 0.667 | 0.000 | - | - | 1.000 |
| 7 | 0.000 | 0.000 | - | - | - |
| 8 | 1.000 | 0.000 | - | - | - |

**Table B6. Distribution of Employees Across Levels (All Education Groups - 698 Employees)**

| Years (Since Entry) | Level 1 | Level 2 | Level 3 | Exit | Total |
|---|---|---|---|---|---|
| 0 | 1.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| 1 | 0.431 | 0.264 | 0.000 | 0.305 | 1.000 |
| 2 | 0.168 | 0.236 | 0.036 | 0.560 | 1.000 |
| 3 | 0.070 | 0.136 | 0.070 | 0.723 | 1.000 |
| 4 | 0.033 | 0.076 | 0.066 | 0.825 | 1.000 |
| 5 | 0.013 | 0.030 | 0.039 | 0.918 | 1.000 |
| 6 | 0.004 | 0.016 | 0.026 | 0.954 | 1.000 |
| 7 | 0.001 | 0.009 | 0.016 | 0.974 | 1.000 |
| 8 | 0.000 | 0.001 | 0.007 | 0.991 | 1.000 |
| 9 | 0.000 | 0.000 | 0.001 | 0.999 | 1.000 |

**Table B7. Hazard Rates of Exit and Promotion by Level (All Education Groups - 698 Employees)**

| Years (at Level) | Level 1 to Exit | Level 1 to Level 2 | Level 2 to Exit | Level 2 to Level 3 | Level 3 to Exit |
|---|---|---|---|---|---|
| 1 | 0.305 | 0.264 | 0.402 | 0.125 | 0.360 |
| 2 | 0.346 | 0.259 | 0.437 | 0.333 | 0.200 |
| 3 | 0.359 | 0.222 | 0.500 | 0.150 | 0.417 |
| 4 | 0.367 | 0.163 | 0.286 | 0.286 | 0.571 |
| 5 | 0.522 | 0.087 | 1.000 | 0.000 | 0.333 |
| 6 | 0.556 | 0.111 | - | - | 0.500 |
| 7 | 0.667 | 0.000 | - | - | 0.000 |
| 8 | 1.000 | 0.000 | - | - | - |
| 9 | - | - | - | - | - |

## B.2 Descriptive Statistics for the Estimation Sample

When restricting attention to individual histories for which no level or performance rating is missing, the largest reduction in the number of observations is due to missing performance ratings. Of the original 21,905 employee-years (2,714 individuals) entering the firm at Level 1 between 1970 and 1979, 20,212 employee-years (2,557 individuals, of which 1,552 have 16 or more years of education at entry) have no level information missing.

Only 1,921 (699 individuals, of which 502 have 16 or more years of education at entry) have no level or performance rating missing. Of these 699, only individual who was assigned to Level 3 in period 2. Of the remaining 698, 43 employees have 12 years of education at entry, 70 individuals have 13 to 15 years and 502 have 16 or more years (for 83 individuals education information is missing at the time of entry).

As for the distribution of employees across levels (Table B8), restricting attention to the individuals with at least 16 years of education at entry, the patterns are very similar across the groups of 1,552 individuals, for which no level information is missing, and the group of 502 individuals, for which no level or performance rating is missing. Still, exit in this latter group is much less pronounced, given that individuals with missing ratings tend to have longer tenures, as reflected in the hazard rate of separation reported in Table B9.

Finally, Table B10 contains the estimated values of the vector of structural parameters of the model and the associated standard errors.

**Table B8. Distribution of Employees Across Levels**
**(16 or More Years of Education at Entry - 502 Employees)**

| Years (Since Entry) | Level 1 | Level 2 | Level 3 | Exit | Total |
|---|---|---|---|---|---|
| 0 | 1.000 | 0.000 | 0.000 | 0.000 | 1.000 |
| 1 | 0.347 | 0.327 | 0.000 | 0.327 | 1.000 |
| 2 | 0.100 | 0.277 | 0.042 | 0.582 | 1.000 |
| 3 | 0.034 | 0.141 | 0.084 | 0.741 | 1.000 |
| 4 | 0.014 | 0.074 | 0.070 | 0.843 | 1.000 |
| 5 | 0.008 | 0.026 | 0.048 | 0.918 | 1.000 |
| 6 | 0.002 | 0.012 | 0.034 | 0.952 | 1.000 |
| 7 | 0.000 | 0.006 | 0.018 | 0.976 | 1.000 |
| 8 | 0.000 | 0.002 | 0.008 | 0.990 | 1.000 |
| 9 | 0.000 | 0.000 | 0.002 | 0.998 | 1.000 |

**Table B9. Hazard Rates of Exit and Promotion by Level**
**(16 or More Years of Education at Entry - 502 Employees)**

| Years at Level | Level 1 to Exit | Level 1 to Level 2 | Level 2 to Exit | Level 2 to Level 3 | Level 3 to Exit |
|---|---|---|---|---|---|
| 1 | 0.327 | 0.327 | 0.390 | 0.128 | 0.429 |
| 2 | 0.368 | 0.345 | 0.468 | 0.329 | 0.273 |
| 3 | 0.280 | 0.380 | 0.563 | 0.188 | 0.375 |
| 4 | 0.353 | 0.235 | 0.250 | 0.500 | 0.600 |
| 5 | 0.286 | 0.143 | 1.000 | 0.000 | 0.500 |
| 6 | 0.500 | 0.250 | - | - | 0.000 |
| 7 | 1.000 | 0.000 | - | - | 0.000 |
| 8 | - | - | - | - | - |
| 9 | - | - | - | - | - |

**Table B10. Parameter Estimates (Standard Deviations in Parenthesis)**

| | | | |
|---|---|---|---|
| $a_\beta$ | 1.000 | $b_\beta$ | 1.000 |
| | (248.791) | | (166.960) |
| $\alpha_1$ | 0.869 | $\underline{y}_1$ | $-2,446.885$ |
| | (0.173) | | (199.442) |
| $\alpha_2$ | 0.778 | $\underline{y}_2$ | $-5,986.493$ |
| | (0.486) | | (362.882) |
| $\alpha_3$ | 0.999 | $\underline{y}_3$ | $-880,226.430$ |
| | (307.169) | | (11,046,539.227) |
| $\beta_1$ | 0.069 | $\xi_3$ | 0.564 |
| | (0.012) | | (166,960.868) |
| $\beta_2$ | 0.000 | $E_1(\overline{\theta})$ | 0.010 |
| | (0.000) | | (0.687) |
| $\beta_3$ | 0.700 | $E_2(\overline{\theta})$ | 0.002 |
| | (2.486) | | (442.729) |
| $\delta$ | 0.950 | $E_3(\overline{\theta})$ | 0.111 |
| | (0.073) | | (24.558) |
| $\overline{y}_1$ | 50.940 | $E_1(\underline{\theta})$ | 0.000 |
| | (282.741) | | (0.249) |
| $\overline{y}_2$ | 3,599.936 | $E_2(\underline{\theta})$ | 0.987 |
| | (19,046.524) | | (0.433) |
| $\overline{y}_3$ | 40,846.745 | $E_3(\underline{\theta})$ | 0.001 |
| | (38,654.301) | | (2.195) |

## References


[1] 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.

[2] Baker, G., M. Gibbs and B. Holmström (1994b): "The Wage Policy of a Firm", *Quarterly Journal of Economics* 109(4), 921-955.

[3] Baker, G., M. Gibbs and B. Holmström (1995): "Internal Labor Markets: Too Many Theories, Too Few Facts", *American Economic Review* 85(2), 255-259.

[4] Banks, J.S., and R.K. Sundaram (1992a): "Denumerable-Armed Bandits", *Econometrica* 60(5), 1071-1096.

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

[6] Chade, H., and E.E. Schlee (2002): "Another Look at the Radner-Stiglitz Nonconcavity in the Value of Information", *Journal of Economic Theory* 107(2), 421-452.

[7] Chiappori, A.P. (2003): "Testing Contract Theory: a Survey of Some Recent Work", in Advances in Economics and Econometrics - Theory and Applications, Eighth World Congress, M. Dewatripont, L. Hansen and P. Turnovsky, ed., Econometric Society Monographs, Cambridge University Press, Cambridge, 2003, 115-149.

[8] Chiappori, A.P., B. Salanié and J. Valentin (1999): "Early Starters Versus Late Beginners", *Journal of Political Economy* 107(4), 731-760.

[9] Eeckstein, Z., and K. Wolpin (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities", *Econometrica* 67(6), 1295-1339.

[10] Fairburn, J.A., and J.M. Malcomson (2001): "Performance Promotion and the Peter Principle", *Review of Economic Studies* 68(1), 45-66.

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

[12] Gibbons, R., and M. Waldman (1999b): "A Theory of Wage and Promotion Dynamics Inside Firms", *Quarterly Journal of Economics* 114(4), 1321-1358.

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

[14] Holmström, B., and P. Milgrom (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics and Organization* 7, 24-52.

[15] Holmström, B., and J. Tirole (1989): "The Theory of the Firm", in *Handbook of Industrial Organization*, Volume I, Chapter 2, R. Schmalensee and R.D. Willig Eds., North-Holland, Amsterdam, the Netherlands, 61-133.

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

[17] Jovanovic, B. (1979b): "Firm-specific Capital and Turnover", *Journal of Political Economy* 87(6), 1246-1260.

[18] Kakigi, R. (1983): "A Note on Discounted Future Two-Armed Bandits", *Annals of Statistics* 11(2), 707-711.

[19] Keane, M.P., and R.M. Sauer (2003): "A Computationally Practical Simulation Estimation Algorithm for Dynamic Panel Data Models with Unobserved Endogenous State Variables", mimeo, Yale University.

[20] Keane, M.P., and K. Wolpin (2001): "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment", *International Economic Review* 45(2), 601-641.

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

[22] Kwon, I. (2004): "Threat of Dismissal: Incentives or Sorting?", University of Michigan, mimeo.

[23] Nagypal, E. (2003): "Learning-by-Doing Versus Learning About Match Quality: Can We Tell Them Apart?", Northwestern University, mimeo.

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

[25] Prendergast, C. (1993): "The Role of Promotion in Inducing Specific Human Capital Acquisition", *Quarterly Journal of Economics* 108(2), 523-534.

[26] Waldman, M. (1984b): "Job Assingment, Signalling, and Efficiency", *Rand Journal of Economics* 15(2), 255-267.