

OPTIMAL EMPLOYEE RETENTION WHEN INFERRING UNKNOWN LEARNING CURVES

Alessandro Arlotto
Noah Gans

University of Pennsylvania
3730 Walnut Street, Suite 500
Philadelphia, PA 19104-6340 U.S.A.

Stephen Chick

INSEAD
Technology and Operations Management Area
Boulevard de Constance
77305 Fontainebleau FRANCE

ABSTRACT

This paper formulates an employer's hiring and retention decisions as an infinite-armed bandit problem and characterizes the structure of optimal hiring and retention policies. We develop approximations that allow us to explicitly calculate these policies and to evaluate their benefit. The solution involves a balance of two types of learning: the learning that reflects the improvement in performance of employees as they gain experience, and the Bayesian learning of employers as they infer properties of employees' abilities to inform the decision of whether to retain or replace employees. Numerical experiments with Monte Carlo simulation suggest that the gains to active screening and monitoring of employees can be substantial.

1 INTRODUCTION

This paper examines the problem of deciding how long to observe workers before deciding to retain or to replace them with new hires. Workers are assumed to learn and improve their performance through time, but the base levels of their abilities are both heterogeneous and unknown. An employer can only infer the abilities of employees by observing their performance over time.

Uncertainty, together with variation across employees, across tasks and over time, makes decisions regarding the retention of workers complex. The longer a worker is retained, the better an inference an employer can make regarding his or her attributes. On-the-job learning, which can lead to quality improvements in incumbent employees, also favors employee retention. Yet the opportunity cost of retaining a poor performer can be great, particularly if there is wide variation in quality across the population of potential hires.

The problem arises in a variety of sectors, including call centers and health care services. We summarize our initial findings from work in progress (Arlotto, Chick, and Gans 2010) that motivates the problem, formulates a mathematical model that is studied in detail, and provides numerical results based on dynamic programming and on Monte Carlo simulation analysis.

Several streams of research address elements of this problem. The literature on secretary problems addresses the initial screening and hiring of employees from a heterogeneous pool (Freeman 1983). Similarly, work on multi-armed bandits that addresses traditional labor-market problems (in which employees choose firms), can be reinterpreted and remodeled to address firms choosing employees (Jovanovic 1979, Banks and Sundaram 1992). These streams do not take worker learning into account, however. Conversely, there is vast empirical literature on learning curve phenomena at the individual, team, and organization levels (Yelle 1979), as well as papers devoted to effective managerial control of factors that affect or depend on learning (Dada and Srikanth 1990, Wiersma 2007). Few of these papers the question of how heterogeneity influences retention decisions.

A few recent papers in scientific literature do address these dimensions of heterogeneity in learning and employee retention. Shafer, Nembhard, and Uzumeri (2001) provide empirical evidence of the heterogeneity of learning curves across individuals who assemble car radios, and Pisano, Bohmer, and Edmondson (2001) document heterogeneity of learning curves across hospital units that perform cardiac surgery. In other works, Mazzola and McCardle (1996) and Mazzola and McCardle (1997) develop models to estimate uncertain learning curves and to control production run lengths, given that a firm faces this uncertainty. Nembhard (2001) proposes a heuristic for assigning workers to tasks, based on differences among individual learning rates. Most papers do not model the uncertainty regarding learning curves across individuals or groups. Nor do they address employee turnover or employee retention decisions.

We develop and analyze a model that integrates all of these factors. In our model, an employer (referred to as "she") seeks to hire and retain a fixed number of employees from an infinite, heterogeneous population of potential hires. Each employee (referred to as "he") repeatedly performs the same task, whose cost the employer wishes to minimize or, equivalently, whose quality is to be maximized. Each hire moves down a learning curve, but elements of

the curve's parameters are unknown to the employer. The employer takes a Bayesian view of employees' types, and by repeatedly observing the outcomes of a given worker's tasks, she can make increasingly better judgments concerning his quality. After each such task, the employee may decide end his employment, turning over. Given the worker decides to stay, the employer can decide whether to retain him or to replace him with a new hire.

We formulate this problem as an infinite-horizon, discounted problem in which the employer uses one worker at each time, and we show that it is essentially an infinite-armed bandit problem. We show that the following results from prior related work by [Banks and Sundaram \(1992\)](#) can be extended to our, more complex, problem.

- The employer can use a worker's prior distribution and tenure to calculate a so-called Gittins index, and at any time it is optimal for the employer to engage a Gittins-index minimal worker.
- It is optimal to retain current employees as long as their Gittins indices compare favorably to those of potential hires.
- If a current employee's Gittins index is not optimal, however, then it is optimal to hire a new worker and to never return to the current employee.

Given the availability of a Gittins index, the above policy is both intuitive and straightforward to execute. Unfortunately, the Gittins index itself is difficult to calculate. Nevertheless, for specific common forms of the learning-curve function and given outcomes that can be appropriately transformed into normally distributed data with known variances and unknown means (with a conjugate prior distribution), we:

- develop approximations to the Gittins index that are straightforward to calculate and implement;
- compute a simple stopping boundary such that, if the posterior mean is below it, then the current employee should be retained;
- demonstrate that the stopping boundary reflects a tradeoff between two types of learning: the employee's performance improvement that is linked to on-the-job experience, and the employer's statistical learning that allows for better judgment concerning the worker's ability; and
- perform numerical experiments which suggest that the value of active monitoring and screening of employees can be substantial and demonstrate how hiring costs and the screening rate of employees should be balanced.

Section 2 formally defines employee behavior and the problem for an employer that requires the services of a single worker. It also summarizes general theoretical results. Section 3 describes the dynamic programming and simulation tools that we use to explicitly calculate the Gittins indices and the optimal stopping boundary for screening employees, assuming a relatively mild additional structure for the type of learning. The simulations are interesting in that they model the Bayesian inference process for the unknown learning parameters of employees as part of employers' hiring and retention decisions. The analysis of Section 4 shows the economic value of actively monitoring and screening of employees. [Arlotto, Chick, and Gans \(2010\)](#) provide a more thorough literature review, a fuller theoretical treatment, mathematical proofs, additional numerical studies, more managerial insights, and extensions to account for hiring costs and for multiple parallel workers.

2 THE HIRING AND RETENTION PROBLEM WITH ONE EMPLOYEE

This section formalizes the problem of an employer who requires the services of a single worker and who, in each discrete period of time, can decide whether to retain the current employee or to terminate him and hire a new one.

2.1 Formal Statement of the Hiring and Retention Problem

This section describes our model of the performance of individual workers, the heterogeneity of their learning curves and the process by which they might end their willingness to be employed. We then describe how employers choose workers sequentially and how employers infer the learning curve of the employees. The employers' problem of hiring and retention will be expressed in terms of minimizing the expected discounted infinite horizon cost of employing a sequence of workers.

We model the performance of potential workers as being uncertain, heterogenous, and evolving over time. We begin by describing the daily performance of worker i on day n through the relation

$$Z_{i,n} = g(a_i, n, \varepsilon_{i,n}), \text{ for all } i \in \mathcal{S}_0 \tag{1}$$

where \mathcal{S}_0 is a countably infinite set of potential workers, a_i reflects worker i 's ability, n reflects his experience, $\varepsilon_{i,n}$ is a noise term that reflects random variations across outcomes, and $g(\cdot)$ is a deterministic function of its parameters.

For example, [Yelle \(1979\)](#) describes the following commonly-used form.

$$Z_{i,n} = \exp(a_i + b_i \ln(n) + \varepsilon_{i,n}) \tag{2}$$

Here, a_i determines a base-level cost or task time, and $b_i < 0$ describes the rate of learning. Given normally-distributed $\varepsilon_{i,n}$, the $Z_{i,n}$'s are log-normally distributed, an assumption that is consistent with characterizations of learning that are found in a number of empirical studies (Brown et al. 2005, Gans et al. 2010). As n approaches infinity, $b_i \ln(n)$ would lead to a performance that converges to zero. Functional forms that lead to stochastic performance with a nonzero mean include $b_i n / (n + \zeta)$ and $b_i \ln(n) / (n + \zeta)$, for some $\zeta > 0$.

Associated with each worker is also an outcome, λ_i , after which he turns over and decides to quit or retire. If the employer were to know each worker's a_i and λ_i , then she would know the evolution of that worker's performance until turnover and could calculate the expected discounted cost or value of the services rendered, $\sum_{j=1}^{\lambda_i} \theta^{j-1} \mathbb{E}[cZ_{i,j}]$, where the one-period cost is a linear function of each outcome, for some factor $c > 0$, and the single-period discount rate is θ . Observe that the discounting scheme assumes that outcomes are evenly spread through time.

But the employer does not know each employee's a_i or λ_i . Rather, she believes that there exists a distribution of abilities in the population of potential workers, A , as well as an analogous distribution of lifetimes, Λ . These distributions can naturally be estimated using historical data and statistical techniques. Each time the employer hires a new worker, she views that worker's A_i and Λ_i as *iid* samples from the population distributions.

The employer requires the services of a single employee drawn from an infinite pool of potential workers, $i \in \mathcal{S}_0$, and at any time $t \in \{0, 1, 2, \dots\}$ she can decide which worker to employ. We let $\pi(t) = i$ denote the employer's choice of employee i at time t . At $t = 0$ all potential workers have the same prior distributions, A_i and Λ_i , so the employer is indifferent among her choices. But with experience the distribution of a given employee's outcomes may evolve. Furthermore, the employer can use worker i 's performance history up to time t , $\mathcal{H}_i^t = \{(z_{i,s}) \mid \pi(s) = i, s = 1, 2, \dots, n_i(t)\}$, where $n_i(t)$ is the number of times that employee i has been used up to time t , to update her beliefs concerning the distributions of the random parameters A_i and Λ_i .

For simplicity we will assume that each A_i is independent of each Λ_i . Thus, while the employer may learn about both A_i and Λ_i , information concerning one is not informative about the other.

We let $v_{i,n_i}(a) = \mathbb{P}\{A_i \leq a \mid \mathcal{H}_i^{n_i}\}$ denote the probability distribution that describes the employer's uncertainty concerning A_i after the n_i th outcome. More specifically, $Z_{i,n_i+1} \equiv Z(v_{i,n_i}, n_i)$, and for $A_i = a$, the performance $\{Z(v_{i,n_i}, n_i) \mid A_i = a\}$ has distribution $\xi_{a,n_i}(z) = \mathbb{P}\{Z(v_{i,n_i}, n_i) \leq z \mid A_i = a\}$. The expected outcome of employing worker i with age n_i and belief distribution v_{i,n_i} is then

$$\mathbb{E}[Z(v_{i,n_i}, n_i)] = \int \int z d\xi_{a,n_i}(z) dv_{i,n_i}(a).$$

In turn, the marginal distribution of performance is a mixture of the ξ_{a,n_i} 's, namely $\xi_{n_i}(z) = \int \xi_{a,n_i}(z) dv_{i,n_i}(a)$.

An untried worker, i , has prior experience $n_i \equiv 0$ and prior distribution $v_{i,0} \equiv v_0$, where v_0 is a common prior distribution of A_i for all potential workers. After the completion of a task, worker's i experience, n_i , increases deterministically by one so that time $t = \sum_{i \in \mathcal{S}_0} n_i$ counts the number of completed tasks, and the employer updates the distribution of her belief concerning i 's ability according to Bayes' rule. Let $\mathcal{P}(\mathbb{R}_+)$ be the set of all probability measures, ν , on the Borel sets \mathcal{B} of \mathbb{R}_+ , and consider the Bayes operator $\beta : \mathcal{P}(\mathbb{R}_+) \times \mathbb{R}_+ \rightarrow \mathcal{P}(\mathbb{R}_+)$ defined as

$$v_{i,n_i+1}(a) = \beta(v_{i,n_i}, z; a) = \frac{v_{i,n_i}(a) d\mathbb{P}\{Z_{i,n_i} \leq z \mid A \leq a\}}{d\mathbb{P}\{Z_{i,n_i} \leq z\}}. \tag{3}$$

Thus for any given observation, z , the Bayes operator maps the prior distribution, v_{i,n_i} to its posterior, v_{i,n_i+1} .

After each outcome, the current employee notifies the employer of his intention to continue working or to quit, and with this information the employer updates her belief concerning his residual lifetime. For employee i we define the *quitting* probability, $q_{i,n}$, to be

$$q_{i,n} = \mathbb{P}\{\Lambda_i = n \mid \Lambda_i \geq n\} \tag{4}$$

and call $1 - q_{i,n}$ i 's *continuation* probability. Suppose that \mathcal{S}_0 is the initial set of potential workers (e.g., $\mathcal{S}_0 = \{1, 2, \dots\}$). At any time t , if employee i quits at time t we remove him from the list of those who are employable: $\mathcal{S}_{t+1} = \mathcal{S}_t \setminus \{i\}$. At time t the index $r_i = \mathbb{1}_{\{i \notin \mathcal{S}_t\}}$ denotes employee's status as having quit ($r_i = 1$) or not ($r_i = 0$).

At each time t the employer chooses $\pi(t) = i \in \mathcal{S}_t$, and a hiring and retention policy, $\pi = \{\pi(0), \pi(1), \dots\}$, defines which workers the employer engages over time. Let Π denote the set of non-anticipating policies, whose decisions require only a knowledge of the performance histories of the workers, $\mathcal{H} = \cup_i \mathcal{H}_i$. For any policy $\pi \in \Pi$, we note that $n_{i,t} = \sum_{s=0}^{t-1} \mathbb{1}_{\{\pi(s)=i\}}$ is the experience of worker i to date at time t .

The expected cost of policy $\pi \in \Pi$ is therefore

$$C_\pi(v_0) = \lim_{T \rightarrow \infty} \mathbb{E} \left[c \sum_{t=0}^T \theta^t Z(\pi(t), n_{\pi(t)}) \right], \tag{5}$$

where $c > 0$ is the cost per unit of performance in a given outcome, and where we refer to what would formally be $Z(v_{\pi(t),n_{\pi(t),t}},n_{\pi(t),t})$ by the simplified notation $Z(\pi(t),n_{\pi(t)})$.

For the problem to be analytically tractable, we assume that outcomes are uniformly bounded. That is, there exist $-\infty < K_{\text{inf}} \leq K_{\text{sup}} < \infty$, such that $K_{\text{inf}} \leq \inf_{v_i,n_i} Z(v_i,n_i) \leq \sup_{v_i,n_i} Z(v_i,n_i) \leq K_{\text{sup}}$ almost surely. This allows us to interchange limits and integrals in (5) to obtain

$$C_{\pi}(v_0) = \mathbb{E} \left[c \sum_{t=0}^{\infty} \theta^t Z(v_{\pi(t)},n_{\pi(t)}) \right]. \tag{6}$$

For the case of an employer that always has one employee, the hiring and retention problem involves finding a policy $\pi^* \in \Pi$ that minimizes the expected discounted cost over all time:

$$\pi^* = \operatorname{argmin}_{\pi \in \Pi} C_{\pi}(v_0). \tag{7}$$

2.2 Summary of Theoretical Analysis of the Hiring and Retention Problem

The employer’s hiring and retention problem is closely related to classic Bayesian bandit problems. Each potential employee’s state is a triple that represents his availability, r_i , his experience, n_i , and the employer’s current belief regarding his ability, v_{i,n_i} . In each period, the employer chooses exactly one employee, and this choice yields a single-period cost, along with a state transition of only that employee; the states of all other potential hires remain unchanged. From (3) – (4) we see that state transitions are Markovian. One difference between the employer’s and traditional bandit problems is that, in the former, an employee becomes unavailable when he quits, while in the latter arms are always available.

Arlotto, Chick, and Gans (2010) derive analytical properties of the hiring and retention problem by first noting that a policy is *stationary* if, at any time t , the action it prescribes in a given state is independent of t , and a policy is *deterministic* if the action it prescribes is never randomized. Then given $Z \in [K_{\text{inf}}, K]$, a finite m , and $\theta \in [0, 1]$, the employer’s problem is an infinite-horizon, discounted Markov Decision Process with uniformly bounded costs, a fact which implies that there exists an optimal hiring and retention policy that is both stationary and deterministic (Bertsekas and Shreve 1996, Prop. 9.8).

This fact also justifies the claim that the optimal value function satisfies the Bellman equation. For notational simplicity we drop the employee index, i , and write the Bellman equation as

$$V(v_n, m, n, r) = \begin{cases} \min\{m, HV(v_n, m, n)\}, & \text{if } r = 0, \\ \min\{m, cK + \theta V(v_n, m, n + 1, 1)\}, & \text{if } r = 1, \end{cases} \tag{8}$$

where m is a “retirement” cost, and

$$HV(v_n, m, n) = c\mathbb{E}[Z(v_n, n)] + \theta(1 - q_n)\mathbb{E}[V(\beta(v_n, Z(v_n, n)), m, n + 1, 0)] + \theta q_n V(v_n, m, n + 1, 1). \tag{9}$$

In words, when an employee is productive ($r = 0$), the expected discounted value of continuing his employment, HV , is the expected cost of the current period plus the expected discounted cost to go. The expected discounted cost to go is determined by conditioning on whether the employee remains productive (which happens with probability $1 - q_n$) or not. When $r = 1$ the employee has become unproductive, and the expected discounted value of continuing his employment is the sum of the expected current period cost of an unproductive worker, cK , plus the associated cost to go, which in this case is deterministic.

Following the approach used in Whittle (1980), at any time the employer has the option of terminating all employment and paying the retirement cost, m . This retirement cost plays a central role in our analysis.

With this formulation we are able to prove (Arlotto, Chick, and Gans 2010) the following results. Some results extend the analysis of Banks and Sundaram (1992), which hold for arms of bandit problems with a finite number of types, to our situation, where the vector (v_n, n, r) represents an arm and which can take on a continuum of values.

- The hiring and retention problem can be modified slightly so that the original and the modified problems have the same optimal solution, and the modified problem is a bandit problem;
- The optimal solution is based on computing an index for each employee, and selecting the employee with the smallest index at each time, so that the problem possesses a so-called Gittins index policy;
- A worker that is not retained at a given time (because an as-yet unused employee is preferred) will never be selected again;
- The independence from one worker to the next means that the stochastic process of costs incurred each period is a renewal process - with a renewal occurring each time a new employee starts;

- The optimal time to hire a new worker, based on the statistics of the current worker, can be solved with an optimal stopping time formulation;
- For each v_n , n and r , $V(v_n, m, n, r)$ is concave, non-decreasing and Lipschitz continuous in m ;
- The Gittins index is given by

$$M(v_n, n, r) = \sup\{m \in \mathbb{R}_+ \mid V(v_n, m, n, r) = m\}. \tag{10}$$

- While it is not true for bandit problems in general, for this problem the Gittins index of an unused arm (potential employee) equals the expected infinite horizon net discounted cost, $C_{\pi^*}(v_0) = M(v_0, 0, 0)$.

3 IMPLEMENTATION

This section describes mild assumptions that help to make the Bellman equation computable, how we compute the optimal value function using a grid approximation for normally distributed observations, and simulation tools we use to assess the benefit of the optimal policy relative to other easily implementable policies.

3.1 Assumptions That Enable Approximations of the Gittins Index and Optimal Stopping Time

It is optimal to employ a worker whose Gittins index is minimal, and it is sufficient to compare the index of the current employee to that of the pool of potential new hires. Equations (8) and (10) indicate that the Gittins index $M(v_n, n, r)$ for a worker with state (v_n, n, r) is defined with respect to a supremum involving the value function $V(v_n, m, n, r)$.

How can we compute the Gittins indices? When $r = 1$, the current employee has become unproductive. Given (8), the employer’s choice is straightforward: either retire with m or pay cK and continue with the same decision. Algebra shows that the Gittins index is the infinite horizon cost of reusing such an employee, $M(v_n, n, 1) = cK/(1 - \theta)$.

When the current employee is productive ($r = 0$) the computation of the Gittins index is more involved. To enable its computation, we make some simplifying assumptions about the functional form of the n th trial of worker i , $Z_{i,n}$, in (1). We assume $g(\cdot)$ is invertible and that

$$g^{-1}(Z_{i,n}) = a_i + h(n) + \varepsilon_{i,n}, \tag{11}$$

where a_i is an unknown base performance level that may vary across workers, $h(n)$ is a known learning function, and $\varepsilon_{i,n}$ is normally distributed noise with mean 0 and known standard deviation σ . We assume that each potential hire’s base level of performance, A_i , is unknown with initial prior, v_0 , that is $N(\mu_0, \sigma_0^2)$. The subscript i is dropped hereafter for clarity.

The form in (11) allows for the posterior distribution of A to be easily characterized. The random variables $X_n = g^{-1}(Z_{i,n}) - h(n)$ are normally distributed with unknown mean A and variance $\sigma^2 + \sigma_0^2$. By standard Bayesian analysis, the posterior distribution, v_n , of A after observing n calls, $\mathbf{z}_n = (z_1, z_2, \dots, z_n)$ is normal with

$$\mathbb{E}[A \mid \mathbf{z}_n] = \frac{\mu_0 \frac{\sigma^2}{\sigma_0^2} + \sum_{k=1}^n (g^{-1}(z_k) - h(k))}{n + \sigma^2/\sigma_0^2} \text{ and } \text{Var}[A \mid \mathbf{z}_n] = \frac{\sigma_0^2 \sigma^2}{\sigma^2 + n\sigma_0^2}.$$

Define $s_0 = \sigma^2/\sigma_0^2$, and let $s = s_0 + n$ where n is the number of samples observed for the single-worker problem. Set $y_s = \mu_0 s_0 + \sum_{k=1}^n (g^{-1}(z_k) - h(k))$ and $w_s = y_s/s$. The posterior distribution v_n of A given \mathbf{z}_n is thus $N(w_s, \sigma^2/s)$. We can therefore describe (v_n, n) by (w_s, s) .

These assumptions are sufficient to guarantee that both the Bellman equation (8) and the Gittins index (10) are monotone in the posterior mean of A , w_s (Arlotto, Chick, and Gans 2010). The monotonicity of the Gittins index with respect to w_s allows us to simplify the implementation of the optimal policy. It implies that, for each $s = s_0 + n$, there is a simple threshold of performance, $b(s)$, such that it is optimal to retain the employee if $w_s < b(s)$ and to terminate the employee if $w_s > b(s)$.

3.2 Trinomial Grid Approximation for Bellman Equation

We now outline how we use the above results to approximate the value function V , the Gittins index M and the stopping boundary b given that (11) applies, the functions g and h are known and finite for finite values of their arguments, the noise ε_n has zero mean and known variance σ^2 , and the prior distribution for A is $N(w_{s_0}, \sigma^2/s_0)$. We use a technique that is common in computational finance – that of approximating a Gaussian process with a related trinomial process on a grid.

We construct the necessary grid of points for the trinomial tree in the (w, s) coordinate system, estimate the terminal conditions using Monte Carlo simulation, perform a backward recursion using a trinomial tree approximation on the

grid of points in order to approximate both V and the optimal boundary for a given value of m , and then search for the value of m that identifies the Gittins index. That process also identifies the optimal stopping boundary for the hiring and retention problem. The optimal stopping region is approximated by examining the set of points in the grid such that one prefers m to continuing, as in (8) and (9). We use the symbol $\hat{\cdot}$ to refer to the approximations to the optimal value function (\hat{V}) and stopping boundary \hat{b} that results from this process.

The code is implemented in Matlab. Requests for the code can be sent to the authors.

Grid and step size. We defined a grid of points, (w_i, s_j) , such that $w_{i+1} - w_i = \Delta w$, $\forall i = -i_{\min}, -i_{\min} + 1, \dots, -1, 0, 1, \dots, i_{\max} - 1$, and $s_j - s_{j-1} = \Delta s$, $\forall j = 1, 2, \dots, j_{\max}$. We choose Δs so that $\kappa = 1/(\Delta s)$ is an integer number of time steps per unit sample, and choose $s_{\max} = s_{j_{\max}}$ so that the number of samples simulated, $(s_{\max} - s_0)$, is a multiple of the mean time to quit (e.g., $\Delta s = 1/30$ and $s_{\max} = s_0 + 12\mathbb{E}[\Lambda]$).

The size of Δw is chosen to allow the trinomial tree to model the correct evolution of the variance of A according to the following logic. At time s , the posterior variance of A is σ^2/s and at time $s+1$ its posterior variance is $\sigma^2/(s+1)$. Following standard Bayesian results on predictive distributions (DeGroot 1970), the variance in the posterior mean of A at $s+1$ given the value of the posterior mean of A at time s is $\sigma^2/s - \sigma^2/(s+1) = \sigma^2/[s(s+1)]$. A trinomial tree with probability p_u of going up Δw and probability p_d of going down Δw in one time step should therefore satisfy $p_u = p_d$, since the posterior mean is a martingale, and $(p_u + p_d)(\Delta w)^2 = \sigma^2/[s(s+1)\kappa]$. Thus $p_u = p_d = (\sigma^2(\Delta s)/(\Delta w)^2 s(s+1))$ for the κ transitions between time s and time $s+1$ (for all s such that $s - s_0$ is an integer). This probability varies as a function of s . Since p_u increases as s decreases and we are interested in $s \geq s_0$, it is maximized when $s = s_0$. By requiring that $p_u \leq p_{\max} < 1/2$ for some fixed p_{\max} (we used 0.475), one can maximize the feasible size of Δw by setting $\Delta w = \sigma/\sqrt{2\kappa s_0(s_0+1)p_{\max}}$, and we do so.

We choose an interior grid point to precisely match the prior mean, $w_0 = w_{s_0}$, to avoid having to interpolate the value of \hat{V} at the prior mean. We let the range of grid points be large enough so that retirement would be preferred for $w_{s_0} = w_{i_{\max}}$ regardless of $s \geq s_0$ and that continuing would be preferred if $w_{s_0} = w_{i_{\min}}$ for all $s \geq s_0$ (in numerical experiments, values for i_{\max} and i_{\min} were easy to find).

Terminal condition. The terminal condition, $\hat{V}(w_i, m, s_{\max}, 0)$ for each valid i , at time s_{\max} , was approximated as the minimum of m and the sum of the expected discounted cost stream associated with the worker continuing to work until he quits, at which time a discounted cost of m is incurred,

$$\hat{V}(w_i, m, s_{\max}, 0) = \min \left\{ m, \mathbb{E} \left[\sum_{t=s_{\max}}^{\Lambda} cZ(w_t, t) + \theta^{\Lambda-s_{\max}} m \mid w_{s_{\max}} = w_i, \Lambda > s_{\max} \right] \right\},$$

the expectation being estimated with Monte Carlo simulation. This is reasonable because an employee that has survived through time s_{\max} is unlikely to be terminated for poor performance at an even later time. We use sensitivity analysis to insure that variability in the Monte Carlo estimates of the terminal condition do not affect the estimates of the stopping boundary and Gittins indices.

Backward recursion. The recursion step involves implementing κ backward steps that account for the change in value of A during the period of one sample and then a cleanup step at each time s that represents an integer number of samples ($s - s_0$ an integer). The κ backward steps use the trinomial tree to model the normally distributed change in the value of the posterior mean with a multinomial distribution (that is generated from κ independent trinomial outcomes). We use this to estimate the term $\mathbb{E}[V(\beta(v_n, Z(v_n, n)), m, n+1, 0)]$ in (9). The ensuing cleanup step accounts for the sampling costs, quitting and potential firing choice in (8) and (9).

Optimal stopping boundary, given m . For any given value of m and for each time s such that $s - s_0$ is an integer, the cleanup step of the recursion also identifies the grid point that corresponds to the smallest value of w_i that is in the “retirement” set, where one prefers to retire with cost m (which models firing the employee) rather than to continue with the employee. We set that value to be $\bar{b}(s) = w_i$, which is an approximation to the optimal stopping boundary. This bounds the optimal stopping boundary of the approximating process, for any given m , to within Δw for values of $s \in \{s_0, s_0 + 1, s_0 + 2, \dots, s_{\max}\}$. Employees clearly above the boundary $\bar{b}(s)$ can be terminated. Employees whose posterior means are below $\bar{b}(s) - \Delta w$ can be retained.

We have tested several smoothing algorithms to smooth the boundary at “steps” in the estimate of the boundary (that is, points where the value of the boundary changes from one time step to the next time step). Spline interpolation for low numbers of cumulated samples and linear interpolation for large numbers of samples have worked well in the sense of providing smooth boundaries without increasing the expected infinite horizon cost in simulation experiments (within Monte Carlo sampling error to a relative error of less than 0.5%).

Search for the Gittins index. If the prior distribution for the A of an untried worker is $N(w_0, \sigma^2/s_0)$, the Gittins index is defined as the largest m such that one is indifferent between continuing or terminating, $\hat{m}_0 = \sup\{m \in \mathbb{R}_+ \mid \hat{V}(w_{s_0}, m, 0, 0) = m\}$. We equivalently search for the minimum value of m such that the difference between m and $\hat{V}(w_{s_0}, m, 0, 0)$ exceeds a very small threshold that accounts for numerical error (10^{-7}). That is done by a binomial search when suitable values of m can be found to be upper and lower bounds on the Gittins index. A suitable upper bound is $m_{\max} = 1 + cK/(1 - \theta)$ when the support of call times is bounded above by K . For lognormal service times, there is no such K . But if one assumes that the expected service times improve with learning, as they do in our examples,

one can use $m_{\max} = 1.1c\mathbb{E}[Z(w_{s_0}, s_0)]/(1 - \theta)$, which is 10% larger than the expected infinite horizon discounted cost incurred if each performance were a first performance of an employee. Since sampling costs are nonnegative in our examples, we can set $m_{\min} = 0$.

To speed up computation, in our experiments we used the heuristic of setting $m_{\min} = 0.9c\mathbb{E}[Z(w_{s_0}, s_0)]$. The binary search reduces the size of the gap between m_{\max} and m_{\min} by 1/2 at each iteration. In practice, we computed the Gittins index within an absolute error of 0.1 to 1.

Optimal stopping boundary for overall problem. For any given m , there is an associated stopping boundary. The one associated with the Gittins index, \hat{m}_0 , of an untried employee is the boundary such that one would prefer to terminate an employee whose posterior mean of A is above that boundary and incur a one-time cost \hat{m}_0 . By construction, \hat{m}_0 is the infinite-horizon cost of a policy that starts by using a new hire from the pool of potential employees, and then behaving optimally. Thus, the stopping boundary for \hat{m}_0 determines the optimal policy for the optimal hiring and retention problem. We denote that policy by $\hat{\pi}^*$.

3.3 Monte Carlo Simulations of Hiring and Retention Policies

Section 3.2 indicates how we approximate the value function and the Gittins index for the hiring and retention problem by using a trinomial tree, rather than the problem's actual normally-distributed data. The approximation applies when the assumptions of Section 3.1 hold. The approximation is better when the grid is finer.

We checked the accuracy of that approximation, to insure that the grid was fine enough. We did so by using Monte Carlo simulation to assess the infinite horizon cost from (6) for the policy $\hat{\pi}^*$ that is optimal, or at least approximately so, as determined by the procedure in Section 3.2. We observed stability in the results and a concordance between the simulated results (which used a Gaussian process) and the DP approximation (which used a trinomial tree to model noise) for the grid sizes mentioned above. This is true for all results presented below.

We also use simulation to assess the expected costs of other stationary policies that are easy to implement in practice (such as the policy to monitor new employees for n days continuously, replace them if the posterior mean of A for that employee exceeds a threshold, and to keep them until they quit if they have not been replaced based on performance during the first n days).

In each case, we can simulate a policy by noticing a few key aspects. One, the hiring and retention problem is an infinite-horizon renewal process with discounted costs. As such, we can adapt standard techniques, such as renewal-reward / regenerative process simulation, to handle discounted expected infinite horizon costs by simulated a series of independent employees. In particular, if the expected (discounted) cost per renewal cycle is $C_i = \sum_{t=0}^{T_i-1} c\theta^t Z_{i,t}$ with renewal time T_i , then the expected discounted infinite horizon cost is $\mathbb{E}[C_i]/(1 - \mathbb{E}[\theta^{T_i}])$. In our problem, C_i is the cost from employee i until time T_i which is the minimum of the quitting time, Λ_i , and the time that the policy π would terminate the employee. Two, the simulation of most reasonable policies can be represented by checking if the posterior mean for the unknown parameter A_i of employee i is below a stopping boundary $b_\pi(n_i)$ that depends upon his experience n_i . These observations lead to the pseudo-code in Table 1 that simulates different hiring and retention policies π that have stopping boundary b_π .

Table 1: Pseudocode for Monte Carlo Simulation of a Hiring and Retention Policy π .

1. For $i = 1, 2, \dots, N$, generate the i th employees' learning curve parameter a_i from a normal (μ_0, σ_0^2) distribution. Generate λ_i , the number of days the employee will work before quitting (unless he is terminated earlier).
 - (a) For $j = 1, 2, \dots$
 - i. Generate $\varepsilon_{i,j}$ independently from a normal $(0, \sigma^2)$ distribution and compute the realized performance $z_{i,j}$. Compute the posterior means $\hat{\mu}_{i,j} = \mathbb{E}[A_i | \mathbf{z}_{i,j}]$.
 - ii. Incur the appropriate discounted performance cost for that observation
 - iii. If the posterior mean rises above the stopping boundary $\hat{\mu}_{i,j} > b_\pi(j)$ (employee should be terminated) or if $j = \lambda_i$ (employee quits after this observation) then recall the value of j , whether the employee was terminated or had quit, and exit the j th loop to continue with the next employee.
 - (b) Compute the discounted costs for this employee and expected discounted time.
2. Report a histogram to describe the fraction of employees that retire and the fraction that are terminated, as a function of the number of days of tenure, as well as an estimate of the expected infinite horizon cost.

4 NUMERICAL RESULTS

In this section we calculate the Gittins index for the pool of new hires, \hat{m}_0 , as well as the associated optimal stopping boundary for several examples. We simulate the cost of optimal retention, and that of other easily implementable policies. We demonstrate that an active hiring and retention policy has significant value for the employer.

The numerical results below correspond to setting $g(z) = e^z$ and $h(n) = b \ln(n)$. This corresponds to (2) with a common learning parameter $b_i = b$, so that $\ln(Z_{i,n}) = A_i + b \ln(n) + \varepsilon_{i,n}$, where $\varepsilon_{i,n} \sim N(0, \sigma^2)$. This model is consistent with empirical studies (Brown et al. 2005, Shen 2003, Shen and Brown 2006). The numerical results also assume that the times to quitting are *iid* and geometrically distributed, so that $q_n = q = 1/\mathbb{E}[\Lambda_i]$.

4.1 Balancing Uncertainty and Learning Effects

The first example is loosely motivated by a call-center. Each $Z_{i,n}$ represents the average duration (in minutes) of the calls that agent i handles in his n th day of tenure. The distribution of initial service-times across agents, A , has $\mu_0 = -0.40$ and variance $\sigma_0^2 = 0.10$, and the variance in the daily average of the service times is $\sigma^2 = 1.00$. This implies an expected service time of untried agents of $\mathbb{E}[Z_{i,0}] = 1.16$. Worker lifetimes satisfy $\mathbb{E}[\Lambda_i] = 100$, which implies $q = 0.01$. The annual discount rate is 10%, so the one-period discount rate is $\theta = 0.9996$ (based on a year of 250 days), and the unit cost is $c = 1$. Learning is deterministic with rate $b = -0.1255$, which implies that the mean call duration decreases by 50.0% after each year of tenure.

Figure 1 displays the stopping boundary associated with the Gittins index for untried employees, \hat{m}_0 , which, in this example, equals 1,358.9. The left panel plots the stopping boundary with respect to the posterior mean of A , and the solid line in the right panel plots the analogous boundary with respect to the posterior mean of Z_n . An employee whose posterior means fall below these boundaries has a Gittins index below \hat{m}_0 and should be retained, and one whose posterior means fall above the boundary should be replaced by a new hire from the pool of untried potential employees.

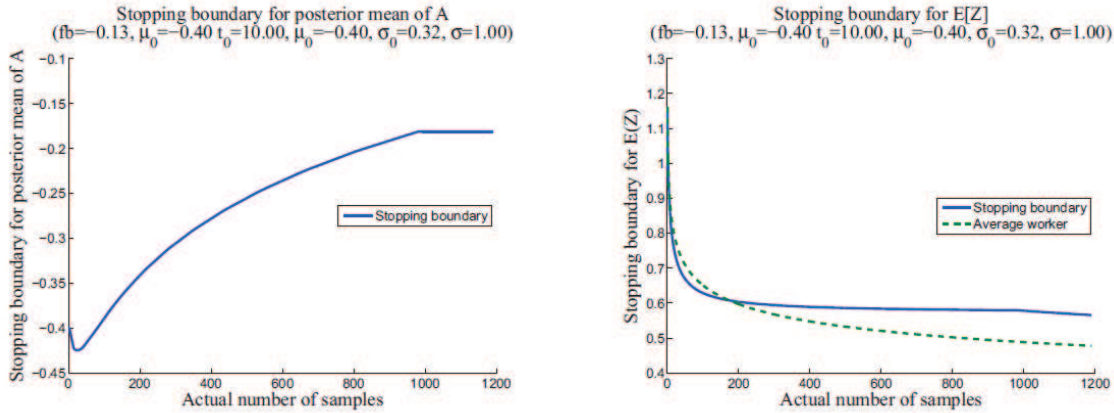


Figure 1: Stopping boundaries for posterior mean of A (left) and Z (right).

In the left panel, we see that the stopping boundary with respect to the posterior mean of A has an interesting “cupped” shape at the left side of the plot, in the early periods of an employee’s tenure. The initial decrease in the acceptable mean reflects the effect of the statistical learning on the part of the employer. As more samples are collected, the uncertainty about the “true” quality of the worker decreases, and the employer is able to screen workers on the basis of a more informative prior. The subsequent climb reflects the gains the employee enjoys as on-the-job experience makes even relatively poor-quality employees attractive candidates for retention. In its right most reaches, the curve appears to increase to an asymptote involving a constant plus $-h(n)$.

The right panel shows the stopping boundary with respect to the mean service time, $\mathbb{E}[Z_n]$. Here, the stopping boundary is monotonically decreasing with a steep initial dip that later flattens out. Unlike the left panel, the right panel does not explicitly display a “dip” that reflects the problem’s two conflicting forces, between the employer’s statistical learning and the employees’ learning by doing. Instead, we find a monotonically decreasing boundary that requires a worker’s expected performance to keep improving over time.

The dashed line in the right panel plots $\mathbb{E}[Z_n]$ for an “average” employee with base-level service time of $A = \mu_0$. The vertical distance between the two curves is a measure of how much better or worse a “marginally retained” employee is in comparison to an “average” employee. During the early periods of an employee’s tenure, only workers whose expected value of Z_n is better than that of the “average” worker are employed, while, after time $t \approx 175$ the stopping boundary allows for retention of workers that have an higher expected value of Z_n .

Table 2 describes how the optimal policy affects employee retention. The policy is particularly selective on the first day of an employee’s tenure, and then it attenuates. In this example, 49.53% of employees are terminated after period one, 28.10% after periods 2 through 10, and 5.17% thereafter. Employee retirement follows the opposite trend. Approximately 5% of the workers resign in the first 20 periods, and the rest later on. Here, 17.20% of employees serve until they naturally turn over and are never terminated. While this fraction of employees terminated seems high, the result assumes there are no hiring costs. Positive hiring costs are reserved for future work.

Table 2: Optimal policy and employee retention.

	Day 1	Days 2–10	Days 11–20	Days 21–∞	Total
Fraction of terminated workers	0.4953	0.2810	0.0318	0.0199	0.8280
Fraction of retired workers	0.0100	0.0257	0.0162	0.1201	0.1720

4.2 How the Optimal Policy Compares with Simpler Policies

This section compares the optimal policy with some other simple policies. The results suggest that employers have good reasons to strictly monitor performance during the early stages of employment.

We consider four alternative hiring policies. In the first, workers are never terminated, and they serve until they naturally turn over. In the second, workers are monitored only on their first day, after which the employer makes a final retention decision. If the worker is retained, he will never be terminated. The third and the fourth policies are similar to the second, but with 10 and 50-day screening periods.

Table 3 summarizes the performance of these policies. The policy that does not monitor employees has an expected-discounted cost that is 24.20% higher than that of the optimal policy. This gap progressively reduces as the monitoring becomes more consistent. The policy that monitors workers during the first 50 days performs nearly as well as the optimal one. This evidence agrees with the observation that most of the termination occurs early on. The fractions of workers that are terminated and that naturally quit change across the various policies in a reasonable manner.

Table 3: Comparison with other hiring policies.

Policy	Total Expected Discounted Cost	Cost increase w.r.t. optimal policy	Fraction of terminated workers	Fraction of retired workers
Optimal policy	1,358.91	-	0.8280	0.1720
Never screen	1,687.71	24.20%	0.0000	1.0000
Screen period 1	1,572.49	15.72%	0.4944	0.5056
Screen periods 1–10	1,398.23	2.89%	0.7758	0.2242
Screen periods 1–50	1,361.12	0.16%	0.8243	0.1757

4.3 How the Optimal Policy Depends on Learning Rates

Section 4.1 studies workers whose performance improves by 50% over 250 days (one year), i.e. $b = -0.1255$. Here, we compare this performance with that of fast-learning workers who improve by 75% in a one-year period ($b = -0.2511$), as well as that of slow-learning workers who improve only by 25% in the same amount of time ($b = -0.0521$). All the other parameters are kept constant, as in Section 4.1.

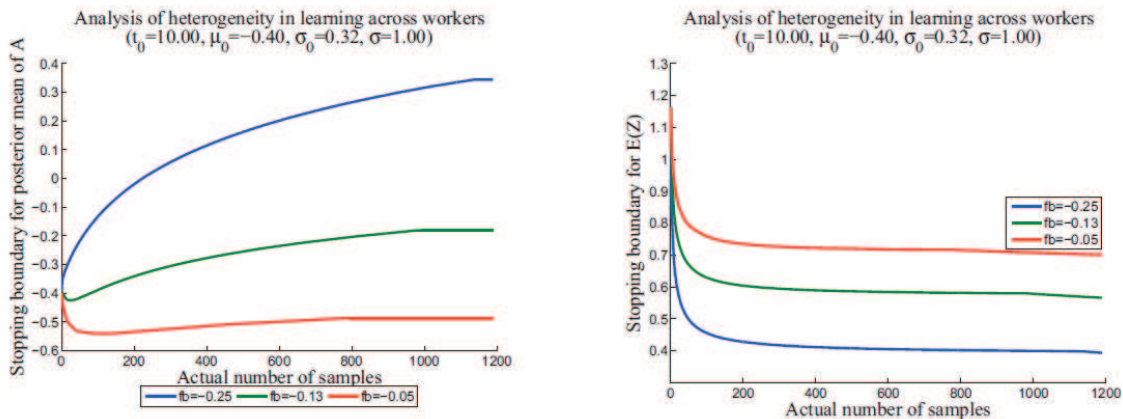


Figure 2: Stopping boundaries for different learning rates

Figure 2 plots the stopping boundary with respect to the posterior mean of A (left) and with respect to $\mathbb{E}[Z_n]$ (right) in these new settings. In the left panel, we notice that the “cupped” shape of the boundary in the early stages of employment is more prominent for the slow learner, and the set of allowable posterior means is smaller. On the other hand, the fast-learning worker immediately benefits of a tangible improvement in his first few days so that the “cupped”

part of the boundary disappears. The contribution of his experience-based learning is so high that the screening policy allows for a large set of posterior means.

A surprising feature is how the stopping boundary with respect to the posterior mean of A translates into the stopping boundary with respect to $\mathbb{E}[Z_n]$. While the ordering of the boundaries may appear to be incorrectly swapped, this is not the case. Rather, fast-learning workers are retained, even though the posterior means of their A_i 's are higher than those of the slower-learning ones. The high learning rates drive their expected mean performance to drop quickly (which is good), and their stopping boundary is the bottom one in the right panel. A similar argument explains why the stopping boundary for the slow learners is the top one in the right panel.

To more clearly understand the effect of changes in employees' learning, we can also look at the values of the Gittins index and at the fraction of terminated workers for these three b 's. Table 4 shows that the optimal retention policy for fast learners has the lowest Gittins index (expected infinite-horizon discounted cost) and the lowest fraction of terminated workers. Slower learners are more expensive and are terminated more frequently.

Table 4: Gittins indexes and termination probabilities with different learning rates.

b	Gittins Index	Fraction of terminated workers			
		Day 1	Days 2–10	Days 11 –20	Total
-0.2511	926.43	0.4065	0.2545	0.0285	0.7028
-0.1255	1,358.91	0.4953	0.2810	0.0318	0.8280
-0.0521	1,686.50	0.5376	0.2978	0.0300	0.8871

5 DISCUSSION AND CONCLUSIONS

This paper models an important problem at the interface of operations research and human resource management: how the collection of information to infer, with Bayes' rule, the abilities of employees interacts with the learning-by-doing that accrues with employees' experience. We made a number of simplifying assumptions. Those assumptions include the absence of hiring costs, the fact that the employer needs one and only one employee at each discrete unit of time, the ability to monitor data about worker performance that directly relates to the employer's financial results, and the ability to terminate and employ with no delay.

Given those assumptions, the work summarized here (and presented in more detail in work in progress) derives a number of interesting results. The problem can be reformulated as an infinite-horizon bandit problem. There is a simple index that can be calculated for employees such that it is optimal to select an employee with the minimal index at every time. The indices satisfy certain monotonicity properties.

Given mild additional assumptions regarding the learning curves of employees, we described an algorithm to compute the indices. The computation of the index for an as-yet untried employee determines an optimal stopping boundary such that, if the mean performance is above that boundary, it is optimal to terminate the employee. Otherwise, retention is optimal. This results in an easy-to-use policy, since the optimal stopping boundary only needs to be assessed once.

While the paper did not focus on this fact, we note that the handling of parameter uncertainty differs in this paper as compared with some of the recent focus in the simulation community. First, the modeling of input uncertainty in this paper is somewhat different than that in other papers. Stochastic uncertainty is included, due to variation in performance each day, even for a given fixed set of parameters. Structural uncertainty is also included, as there is uncertainty about worker performance. Correct modeling of structural uncertainty in this case is not handled as in the usual Bayesian model average, however, as the unknown parameters that describe the population distribution of unknown performances is sampled once per employee. An interesting variation is that the employer is to infer the unknown parameter during the simulation itself.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 0800645. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

REFERENCES

- Arlotto, A., S. E. Chick, and N. Gans. 2010. Hiring and retention of heterogeneous workers. *Working paper - Wharton/INSEAD Alliance*.
- Banks, J. S., and R. K. Sundaram. 1992. Denumerable-armed bandits. *Econometrica* 60 (5): 1071–1096.
- Bertsekas, D. P., and S. E. Shreve. 1996. *Stochastic optimal control: The discrete time case*. Belmont: Athena Scientific.

- Brown, L., N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao. 2005. Statistical analysis of a telephone call center: a queueing-science perspective. *Journal of the American Statistical Association* 100 (469): 36–50.
- Dada, M., and K. N. Srikanth. 1990. Monopolistic Pricing and the Learning Curve: An Algorithmic Approach. *Operations Research* 38 (4): 656–666.
- DeGroot, M. H. 1970. *Optimal statistical decisions*. New York: McGraw-Hill Book Co.
- Freeman, P. R. 1983. The secretary problem and its extensions: a review. *International Statistics Review* 51 (2): 189–206.
- Gans, N., N. Liu, A. Mandelbaum, H. Shen, and H. Ye. 2010. Service times in call centers: Agent heterogeneity and learning with some operational consequences. In *Festschrift in honor of Professor Lawrence D. Brown*, Inst. Math. Stat. Collect., to appear.
- Jovanovic, B. 1979. Job matching and the theory of turnover. *Journal of Political Economy* 87 (5): 972.
- Mazzola, J. B., and K. F. McCardle. 1996. A Bayesian approach to managing learning-curve uncertainty. *Management Science* 42 (5): 680–692.
- Mazzola, J. B., and K. F. McCardle. 1997. The stochastic learning curve: Optimal production in the presence of learning-curve uncertainty. *Operations Research* 45 (3): 440–450.
- Nembhard, D. A. 2001. Heuristic approach for assigning workers to tasks based on individual learning rates. *International Journal of Production Research* 39:1955–1968(14).
- Pisano, G. P., R. M. Bohmer, and A. C. Edmondson. 2001. Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Science* 47 (6): 752–768.
- Shafer, S. M., D. A. Nembhard, and M. V. Uzumeri. 2001. The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Science* 47 (12): 1639–1653.
- Shen, H. 2003. *Estimation, confidence intervals and nonparametric regression for problems involving lognormal distributions*. University of Pennsylvania.
- Shen, H., and L. D. Brown. 2006. Non-parametric modelling for time-varying customer service time at a bank call centre. *Applied Stochastic Models in Business and Industry* 22 (3): 297–311.
- Whittle, P. 1980. Multi-armed bandits and the Gittins index. *Journal of the Royal Statistics Society Series B* 42 (2): 143–149.
- Wiersma, E. 2007. Conditions that shape the learning curve: Factors that increase the ability and opportunity to learn. *Management Science* 53 (12): 1903–1915.
- Yelle, L. E. 1979. The learning curve: historical review and comprehensive survey. *Decision Sciences* 10 (2): 302–328.

AUTHOR BIOGRAPHIES

ALESSANDRO ARLOTTO is a Ph.D. Candidate in the Operations and Information Management (OPIM) Department at the Wharton School of the University of Pennsylvania. His research lies in the realm of applied probability. His web page is opim.wharton.upenn.edu/alear/.

STEPHEN E. CHICK is a professor of Technology and Operations Management at INSEAD, and the Novartis Chair of Healthcare Management. He has worked in the automotive and software sectors prior to joining academia and now teaches operations with applications in manufacturing and services, particularly the health care sector. He enjoys Bayesian statistics, stochastic models, and simulation. His web page is faculty.insead.edu/chick/.

NOAH GANS is a Professor in the OPIM Department at the Wharton School. He worked in business consulting before entering academia and now teaches service operations management. He is interested in call center operations and enjoys stochastic models and applied probability. His email address is gans@wharton.upenn.edu.