

## UPDATE ON ECONOMIC APPROACH TO SIMULATION SELECTION PROBLEMS

Stephen E. Chick

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

Noah Gans

OPIIM Department – Wharton School  
University of Pennsylvania  
3730 Walnut Street, Suite 500  
Philadelphia, PA 19104-6340, U.S.A.

### ABSTRACT

This paper summarizes new analytical and empirical results for the economic approach to simulation selection problems that we introduced two years ago. The approach seeks to help managers to maximize the expected net present value (NPV) of system design decisions that are informed by simulation. It considers the time value of money, the cost of simulation sampling, and the time and cost of developing simulation tools. This economic approach to decision making with simulation is therefore an alternative to the statistical guarantees or probabilistic convergence results of other commonly-used approaches to simulation optimization. Empirical results are promising. This paper also retracts a claim that was made regarding the existence of Gittins' indices for these problems – their existence remains an open question.

### 1 INTRODUCTION

Selecting the best of a finite set of simulated alternatives is a common goal in simulation. There is a great deal of literature in the area of simulation optimization that attempts to address that goal. Much of that literature proposes statistical sampling procedures that provide probability of correct selection guarantees (such as a 95% probability that the correct system is selected, assuming the best is at least \$10K better than the next best, see [Kim and Nelson 2006](#) for an overview), or asymptotic convergence guarantees (such as ensuring the best system is identified with probability one, assuming an infinite number of replications, see [Fu, Glover, and April 2005](#) for an overview). These sampling procedures can be useful for optimizing for a wide variety of metrics, as long as the objective is to maximize or minimize the expected value of the simulation output.

A very different approach to the problem of selecting the best of a finite set of simulated alternatives was presented in [Chick and Gans \(2006\)](#). That approach assumes that managers are concerned about the expected net present

value (NPV) of their decisions. We presume that either the simulation output is itself a measure of the economic merits of the alternative or that the output can be converted to an implied NPV. The manager is motivated to simulate more to reduce uncertainty about the expected NPV of each alternative system, but the manager is motivated to simulate less to avoid the costs of running simulations, as well as to avoid the effect of discounting the expected NPV due to analysis delays.

Before the simulation is built, the manager must decide whether or not to invest in simulation at all. The key issue is whether the simulations will bring enough clarity about which alternative is best, and what expected NPV it will likely bring, to justify the investment in time and money that is required to develop the simulation tool.

Once a simulation tool that can simulate  $k$  different alternative systems is built, the decisions include which system or systems to simulate, for how long, and which alternative to select for implementation. Since the manager is concerned with the expected NPV of her decisions. The expected reward of the ability to simulate each alternative is an input to the decision of whether or not to develop a simulation tool in the first place.

Our approach treats the ability to simulate as a real option, where the alternatives include either simulating, to obtain more information, or stopping and implementing one of the simulated alternatives. We frame the problem in the context of dynamic programming context and seek to provide economically-justified answers to the following managerial questions:

- Should a manager invest the time and money that is required to develop simulation tools?
- If so, for how long should the simulation analysis continue, and which systems should be simulated before stopping to implement an alternative?

This framework therefore links two distinct areas of simulation: (1) The simulation optimization literature, which

presumes that simulation tools already exist, and which focuses on the second question, and (2) the literature on good modeling practice (as in Law and Kelton 2000, §1.7), which assumes that the answer to the first question is yes, and describes how to develop the tools effectively, but does not link the choice to simulate to the economic value that simulation can bring to the firm that uses simulation.

Our formulation is Bayesian: we assume that the manager has prior beliefs concerning the distribution of the NPV of each of the alternatives and that she uses simulation output to update these beliefs. The system which the manager ultimately chooses to implement maximizes expected NPV with respect to the posterior distributions of her beliefs, as well as analysis costs and discounting costs.

Chick and Gans (2006) summarized the problem formulation and outlined how the problem, when there is only one alternative that is assessed with simulation, can be solved using a dynamic programming formulation and an optimal stopping time for a Brownian motion. It also claimed that, when there is more than one alternative that is being assessed with simulation, that a special solution structure called a ‘Gittins index’ can be used to solve the second managerial question. Since the writing of Chick and Gans (2006), we determined a subtle lapse in our proof of the Gittins’ index result. We can neither prove nor disprove the existence of a Gittins-index result at present, so the existence of such a policy is an open question. In Chick and Gans (2008), we construct a simple counterexample that shows that the few existing and relevant results that would guarantee the existence of a Gittins’ index result do not apply.

Nonetheless, Chick and Gans (2008) also provide alternative methods to address the second of the managerial questions above by extending procedures that minimize the (undiscounted) expected opportunity cost selection (Chick and Inoue 2001) to the current context. Further, Chick and Gans (2008) indicate how to approach the first question, given the answer to this second question.

This paper recalls the economic framework to the simulation selection problem in Section 2 and presents a subset of our recent work on this problem.

## 2 PROBLEM DESCRIPTION

A manager seeks to develop one of  $k$  projects, labelled  $i = 1, \dots, k$ . The net present value (NPV) of each of the  $k$  projects is not known with certainty, however. The manager wishes to develop the project which maximizes her expected NPV, or to do nothing if the expected present value of all projects is negative. We represent the “do nothing” option as  $i = 0$  with a sure NPV of zero.

### 2.1 Uncertain Project NPV’s

Let  $X_i$  be the random variable representing the NPV of project  $i$ , where  $X_0 \equiv 0$ . If the manager is risk neutral and the distributions of all  $X_i$ ’s are known to her, then she will select the project with the largest expected NPV,  $i^* = \arg \max_i \{E[X_i]\}$ . Although we model NPVs as simple random variables, the systems that generate them may be quite complex in practice.

It may also be the case that the distributions of the  $X_i$ ’s are not known with certainty by the manager. Rather, she may believe that a given  $X_i$  may come from one of a family of probability distributions,  $P_{X_i|\theta_i}$ , indexed by parameter  $\theta_i$ . We model her belief as taking the form of a probability distribution on  $\theta_i$ , which we call  $P_{\Theta_i}$ . For example, the manager may believe that  $X_i$  is normally distributed with a known variance,  $\sigma_i^2$ , but unknown mean. Then  $P_{\Theta_i}$  represents a probability distribution for the mean. To ease notation, we will sometimes refer to the distribution as  $\Theta_i$ . In this case, the expected NPV of project  $i > 0$  is  $E[X_i] = E[X(\Theta_i)] \triangleq \iint X(\theta_i) dP_{X_i|\theta_i} dP_{\Theta_i}$ . We denote the vector of distributions for the projects by  $\vec{\Theta} = (\Theta_1, \dots, \Theta_k)$ .

### 2.2 Simulation to Select the Best Project

If the distributions of the  $X_i$ ’s are not known, then the manager may be able to use simulation as a tool to reduce distributional uncertainty, before having to decide which project to develop. She may decide to simulate the outcome of project  $i$  a number of times, and she views the result of each run as a sample of  $X_i$ . She uses Bayes’ rule to update her beliefs concerning  $\Theta_i$ .

We model the running of simulations as occurring at sequence of discrete stages  $t = 0, 1, 2, \dots$ , and we represent Bayesian updating of prior beliefs and sample outcomes,  $\{(\vec{\Theta}_t, \vec{X}_t) | t = 0, 1, \dots\}$  as follows. If project  $i > 0$  is simulated at stage  $t$  with sample outcome  $x_{i,t}$ , then  $X_{i,t} = x_{i,t}$  and  $X_{j,t} = 0$  for all  $j \neq i$ . In turn, Bayes’ rule is used to determine  $\vec{\Theta}_{t+1}$ :

$$dP_{\Theta_{i,t+1}}(\theta_i | x_{i,t}, \Theta_{i,t}) = \frac{dP_{X_i|\theta_i}(x_{i,t} | \theta_i) dP_{\Theta_{i,t}}(\theta_i)}{\int_{\theta_i} dP_{X_i|\theta_i}(x_{i,t} | \theta_i) dP_{\Theta_{i,t}}(\theta_i)}$$

for  $\theta_i \in \Omega_{\Theta_i}$ , while  $\Theta_{j,t+1} = \Theta_{j,t}$  for all  $j \neq i$ . So the evolution of the manager’s beliefs regarding the distribution of outcomes of each project is Markovian. We also assume that simulation results, hence the evolution of the manager’s beliefs, are independent from one project to the next.

If, in theory, simulation runs could be performed at zero cost and in no time, then the manager might simulate each of the  $k$  systems infinitely, until all uncertainty regarding the  $\theta_i$ ’s was resolved. At this point the problem would revert

to the original case in which the distributions and means of the  $X_i$  are known.

But simulation runs do take time and cost money. We assume that each run of system  $i$  costs  $\$c_i$  and takes  $\eta_i$  units of time to complete. Thus, given a continuous-time discount rate of  $\delta > 0$ , the decision to simulate system  $i$  costs the manager  $c_i$  plus a reduction of  $\Delta_i = \int_0^{\eta_i} e^{-\delta s} ds < 1$  times the expected NPV of the (unknown) project that is eventually chosen.

There may also be associated up-front costs associated with the development of the simulation tool, itself. For example it may cost time and money to develop the underlying simulation platform, independent of which projects end up being evaluated. Additional costs may be required to be able to simulate particular projects. Furthermore, these project-specific costs may be inter-related.

For the moment, we make two simplifying assumptions regarding the costs of simulation. First, we ignore all up-front costs for the simulation tool, assuming that the necessary facilities exist to simulate all  $k$  projects. Second, we assume that  $\eta_i \equiv \eta$  for all  $k$  projects. This allows us to define a common  $\Delta \equiv \Delta_i$  for the projects as well. Section 4 will show how the first assumption might be relaxed. Chick and Gans (2008) relaxes the second assumption, with some loss of optimality.

Even with these simplifications, the availability of a simulation tool to sample project outcomes makes the manager's problem much more complex. Rather than simply choosing the project that maximizes expected NPV, she must choose a sequence of simulation runs and, ultimately select a project, so that the discounted stream of costs and terminal expected value, together, maximize expected NPV.

We define a number of indices in order to track the manager's choices as they proceed. Let  $T \in \{t = 0, 1, 2, \dots\}$  be the stage at which the manager selects a system to implement. For  $t < T$ , define  $i(t) \in \{1, \dots, k\}$  to be the index of the project simulated at time  $t$ , and define  $I(T) \in \{0, \dots, k\}$  to be the ultimate choice of project.

A *selection policy* is the choice of a sequence of simulation runs, a stopping time, and a final project. Define  $\Pi$  to be the set of all *non-anticipating* selection policies, whose choice at time  $t = 0, 1, \dots$  depends only on system history up to  $t$ :  $\{\vec{\Theta}_0, \vec{X}_0, \dots, \vec{\Theta}_{t-1}, \vec{X}_{t-1}, \vec{\Theta}_t\}$ . Given prior distributions  $\vec{\Theta} = (\Theta_1, \dots, \Theta_k)$  and a policy  $\pi \in \Pi$ , the expected discounted value of the future stream of rewards is

$$V^\pi(\vec{\Theta}) = E_\pi \left[ \sum_{t=0}^{T-1} -\Delta^t c_{i(t)} + \Delta^T X_{I(T),T} \mid \vec{\Theta}_0 = \vec{\Theta} \right], \quad (1)$$

where  $X_{I(T),T}$  is the unknown NPV of the selected system,  $I(T)$ , when a system is selected (at time  $T$ ).

Formally, we define the manager's *simulation selection problem* to be the choice of a selection policy  $\pi^* \in \Pi$  that maximizes  $V^{\pi^*}(\vec{\Theta}) = \sup_{\pi \in \Pi} V^\pi(\vec{\Theta})$ .

### 3 OPTIMAL SIMULATION SELECTION POLICY

Given relatively mild technical conditions, the optimal selection policy  $\pi^*$  to the simulation selection problem in (1) is known to exist, to be stationary, and to stop almost surely at a time  $T$  such that  $E[X_{I(T),T}] \geq \max_{i=1,2,\dots,k} -c_i / (1 - \Delta)$ .

Chick and Gans (2006) noted that the simulation selection problem is what Glazebrook (1979) calls a stoppable family of alternative bandit processes. Glazebrook (1979) states one of the few known results that indicate when a stoppable family of alternative bandit processes has an optimal policy that is an index policy. An index policy is a policy that would, at each step: (a) assign a value to each alternative that depends only on that alternative, and not on the other alternatives, (b) pick the alternative with the highest value and implement the optimal action for that alternative. In the context of simulation selection, the action would be to either simulate that alternative, or to stop simulation in order to implement that alternative. Optimal index policies are also called Gittins-index policies. We originally thought that the optimal policy for the simulation selection problem was a Gittins-index policy, and we developed an asymptotic approximation that could approximate the relevant indices. We have since found an error in the proposed proof and have identified a simple example that shows that Glazebrook's sufficient condition does not apply. Therefore, the question of the existence of a Gittins-index result for (1) remains an open question.

Our asymptotic approximation remains a valid approximation to the optimal expected discounted reward for the simulation selection problem when there is a single simulated system, or when there is a comparison between a single system with an unknown mean NPV and a single alternative with a known NPV.

We have found a nearly optimal algorithm, when there are  $k > 1$  systems, that does not require a Gittins-index result. This is summarized below in this paper, as are numerical results for the special case of jointly independent and normally distributed simulation output with known variances but unknown means. It will be convenient to introduce the variables  $m$  and  $\gamma$  so that

$$\begin{aligned} m &= \text{known NPV of a known alternative, and} \\ -c_i/\gamma &= \text{continuous time approximation to discounted NPV of simulating system } i \text{ forever.} \end{aligned}$$

Further details, examples and theoretical results can be found in Chick and Gans (2008).

### 4 SHOULD I DEVELOP A SIMULATION TOOL?

Suppose that a simulation platform can be developed with a monetary cost of  $\$g_0$  over  $u_0$  units of time. Further suppose

that, once developed, each of  $k$  alternatives can be simulated on this platform. This corresponds to the different system designs being specified by different inputs to the simulation platform. The choice of whether or not to implement the simulation tools depends upon the cost and development time of the tools, as well as the expected reward from selecting a system based upon the simulation output. This expected reward is a function of the simulation selection policy.

The expected discounted reward for the optimal simulation selection policy, when  $k = 1$ , can be approximated by solving a heat equation with a free boundary (Chick and Gans 2008). When  $k > 1$ , however, the expected discounted reward for the optimal policy is not known.

While we cannot explicitly assess the expected value of the optimal simulation selection policy, we know that, by definition, it is at least as large as that of any other policy, including policies that allocate a fixed number of samples in one stage of sampling. In fact, we can easily develop bounds for the optimal expected discounted reward (OEDR) of one-stage policies. Therefore, in a setting in which we want to decide whether or not it is economical to develop simulation tools at all, the economic value of a one-stage allocation policy can be used to evaluate the optimal policy: if the one-stage allocation policy is valuable, then an optimal allocation will be as well. This subsection describes how we evaluate the economic benefit of using one-stage policy, as well as how we use one-stage policies to decide whether or not to build a simulation tool.

Formally, a one-stage allocation  $\mathbf{r} = (r_1, r_2, \dots, r_k)$  maps a given sampling budget of  $\beta \geq 0$  replications to the  $k$  systems, with a total of  $r_i = r_i(\beta) \geq 0$  replications to be run for alternative  $i$ , so that  $\sum_{i=1}^k r_i = \beta$ . For example, the equal allocation sets  $r_i = \beta/k$  (relaxing the integer constraint if needed). After observing those samples, the one-stage allocation policy selects the alternative with the largest posterior expected reward, if that reward exceeds

$$\mu_{00} \triangleq \max\{m, -c_i/\gamma : i = 1, 2, \dots, k\}, \quad (2)$$

and otherwise selects the alternative that maximizes the right hand side of (2). (Recall that  $-c_i/\gamma$  corresponds to simulating alternative  $i$  forever and that  $m$  is the NPV of a known alternative, such as ‘doing nothing’ for  $m = 0$ .)

Suppose further that samples are normally distributed with known variance  $\sigma_i^2$ , but unknown mean whose distribution is Normal  $(\mu_{0i}, \sigma_i^2/t_{0i})$ , with  $\mu_{0i} = y_{0i}/t_{0i}$ . Then the posterior mean that will be realized after the future sampling is performed is the random variable (de Groot 1970)

$$\mathcal{Z}_i \sim \text{Normal} \left( \mu_{0i}, \frac{\sigma_i^2 r_i}{t_{0i}(t_{0i} + r_i)} \right). \quad (3)$$

If we consider the allocation to be a function of  $\beta$  and vary  $\beta$  over all possible allocations, we obtain the following lower bound.

**Lemma 1.** Let  $V^{\pi^*}(\vec{\Theta})$  maximize (1) and  $\mathbf{r}$  be a one-stage allocation. Let  $\mathcal{Z}_i$  be the (random) posterior mean given that  $r_i$  replications for system  $i$  will be run. Then  $V^{\pi^*}(\vec{\Theta}) \geq \underline{\text{OEDR}}(\vec{\Theta})$ , where

$$\underline{\text{OEDR}}(\vec{\Theta}) \triangleq \sup_{\beta \geq 0} e^{-\gamma\beta} \mathbb{E}[\max\{\mu_{00}, \mathcal{Z}_1, \mathcal{Z}_2, \dots, \mathcal{Z}_k\}] - \sum_{i=1}^k r_i c_i. \quad (4)$$

The expectation on the right hand side of (4), in turn, has some easy-to-compute bounds. The bound refers to the order statistics ( $i$ ) for  $i = 0, 1, \dots, k$  such that  $\mu_{0(0)} \leq \mu_{0(1)} \leq \dots \leq \mu_{0(k)}$ .

**Lemma 2.** Let  $\mathbf{r}$  be a one-stage allocation, assume that output is jointly independent and normally distributed with a known variance, and normal prior distribution so that (3) is valid for each  $i$ , let  $\Psi[s] = \int_s^\infty (\xi - s)\phi(\xi)d\xi = \phi(s) - s(1 - \Phi(s))$  be the News vendor loss function for a standard normal distribution, let  $\sigma_{\mathcal{Z},0}^2 = 0$ ,  $\sigma_{\mathcal{Z},i}^2 = \frac{\sigma_i^2 r_i}{t_{0i}(t_{0i} + r_i)}$ , and  $\sigma_{\mathcal{Z},i,(k)}^2 = \sigma_{\mathcal{Z},i}^2 + \sigma_{\mathcal{Z},(k)}^2$ . Then  $\mathbb{E}[\max\{\mu_{00}, \mathcal{Z}_1, \mathcal{Z}_2, \dots, \mathcal{Z}_k\}]$

$$\geq \mu_{0(k)} + \max_{i:i \neq (k)} \sigma_{\mathcal{Z},i,(k)} \Psi \left[ \frac{\mu_{0(k)} - \mu_{0i}}{\sigma_{\mathcal{Z},i,(k)}} \right] \quad (5)$$

$$\mathbb{E}[\max\{\mu_{00}, \mathcal{Z}_1, \mathcal{Z}_2, \dots, \mathcal{Z}_k\}]$$

$$\leq \mu_{0(k)} + \sum_{i:i \neq (k)} \sigma_{\mathcal{Z},i,(k)} \Psi \left[ \frac{\mu_{0(k)} - \mu_{0i}}{\sigma_{\mathcal{Z},i,(k)}} \right]. \quad (6)$$

With perfect information and no discounting or sampling costs, the expected reward of  $\mathbf{r}$  is

$$\overline{\text{OEDR}}(\vec{\Theta}) \triangleq \mathbb{E}[\max\{\mu_{00}, \mathcal{Z}_1, \mathcal{Z}_2, \dots, \mathcal{Z}_k\}]. \quad (7)$$

If  $-g_0 + e^{-\gamma u_0} \overline{\text{OEDR}}(\vec{\Theta}) > \mu_{0(k)}$ , then it would be optimal to invest in the simulation tools that are required to simulate the  $k$  alternatives in question and to evaluate those alternatives, before selecting a project (including the 0 arm). In this case the expected reward from developing the simulation tool and using the allocation  $r_i(\beta)$ , with the choice of  $\beta$  that determined  $\underline{\text{OEDR}}(\vec{\Theta})$ , would exceed 0.

If  $-g_0 + e^{-\gamma u_0} \overline{\text{OEDR}}(\vec{\Theta}) < \mu_{0(k)}$ , however, it would be better to not implement the simulation tools. In this case, even a simulator that could run replications infinitely fast

at no cost would not provide enough information about the system performance, in expectation, to compensate for the time and cost of developing the simulation tools. Rather one should immediately implement project ( $k$ ).

### 5 STOPPING RULES BASED ON ECONOMICS

We now presume that a platform used to simulate  $k$  systems has been built. We seek an effective sequential simulation selection policy and guidance for when to stop simulating in favor of implementing a system. We appeal to another class of one-stage policies to determine whether or not it is valuable to continue simulating, and if so, which alternative to simulate. Those one-stage policies seek to maximize the expected (undiscounted) reward over a finite horizon.

In particular, we examine the one-stage  $\mathcal{L}\mathcal{L}$  allocation (Chick and Inoue 2001), which minimizes a bound on the expected opportunity cost (EOC) of a potentially incorrect selection. Gupta and Miescke (1996) show that minimizing the EOC is equivalent to maximizing the posterior mean that is realized once a finite total number of samples is observed. Thus, the  $\mathcal{L}\mathcal{L}$  allocation seeks to maximize the expected undiscounted reward over a finite horizon.

Section 6 adapts and extends the  $\mathcal{L}\mathcal{L}$  allocation to the current context, in which both discounting and sampling costs are included. The resulting sequential procedure assumes that samples are normally distributed with a known variance that may differ for each alternative.

The general idea of our sequential sampling procedure is simple. At each stage of sampling, the procedure first tests whether or not to continue sampling. It does this by checking if there exists some one-stage  $\mathcal{L}\mathcal{L}$  allocation of  $\beta$  samples, for some  $\beta \geq 1$ , that leads to an expected discounted reward that exceeds the value of stopping immediately. If there is value to continuing, then one replication is run for the alternative that  $\mathcal{L}\mathcal{L}$  suggests would most warrant an additional replication. After that replication is run, the statistics for that system are updated, with the posterior distribution from the current stage becoming the prior distribution for the next stage. If there is no value to continuing for any  $\beta \geq 1$ , then the procedure stops.

Before presenting the sequential sampling procedure, we focus on stopping rules for the scheme. The development of Section 4 immediately suggests a mechanism to formalize whether or not there is value to additional sampling. One should continue to sample if  $\text{OEDR}(\tilde{\Theta}) > \mu_{0(k)}$ . This will happen if there is a one-stage allocation of size  $\beta$  that leads to value for continuing to simulate. Unfortunately, the sequential recalculation of  $\text{OEDR}(\tilde{\Theta})$  that would be required by such a procedure is computationally burdensome. Fortunately, there is an easy-to-compute alternative. Substituting the right hand side of (5) for the expectation in the right hand side of (4) leads to an easily computable and analytically justifiable bound.

**Stopping rule  $\text{EOC}_1^Y$**  (with implicit one-stage allocation  $r_i = r_i(\beta) \geq 0$  such that  $\sum_{i=1}^k r_i = \beta$ ): Continue sampling if and only if there is a budget  $\beta \geq 1$  such that

$$\frac{\mu_{0(k)} + \max_{i:i \neq (k)} \left\{ \sigma_{\mathcal{L},i,(k)} \Psi \left[ \frac{\mu_{0(k)} - \mu_{0i}}{\sigma_{\mathcal{L},i,(k)}} \right] \right\}}{e^{\gamma\beta}} - \sum_{i=1}^k r_i c_i \quad (8)$$

$$> \mu_{0(k)}.$$

In practice,  $\text{EOC}_1^Y$  may not be as effective as hoped. It may sample less than is optimal because  $\text{EOC}_1^Y$  accounts for only a subset of the economic value of sampling. In numerical experiments, the expected discounted reward is greater if one samples somewhat more than is optimal, as compared to sampling somewhat less than is optimal. The next stopping rule, which may be less justifiable analytically, increases sampling slightly by plugging the right hand side of the upper bound in (6) into the expectation of (4).

**Stopping rule  $\text{EOC}_k^Y$**  Continue sampling if and only if there is a budget  $\beta \geq 1$  such that

$$\frac{\mu_{0(k)} + \sum_{i:i \neq (k)} \sigma_{\mathcal{L},i,(k)} \Psi \left[ \frac{\mu_{0(k)} - \mu_{0i}}{\sigma_{\mathcal{L},i,(k)}} \right]}{e^{\gamma\beta}} - \sum_{i=1}^k r_i c_i > \mu_{0(k)}. \quad (9)$$

Section 6 fully specifies how these stopping rules are used with the  $\mathcal{L}\mathcal{L}$  allocation to solve the simulation selection problem.

### 6 A SIMULATION SELECTION PROCEDURE

We now adapt and extend the  $\mathcal{L}\mathcal{L}$  allocation to the current context, in which both discounting and sampling costs are included. The resulting sequential procedure assumes that samples are jointly independent and normally distributed with a known variance that may differ for each alternative. We leave unknown sampling variances for future work.

The one-stage  $\mathcal{L}\mathcal{L}$  allocation allocates a finite number of samples to  $k$  alternatives in a way that maximizes the expected (undiscounted) reward at the end of sampling. Because the optimal solution is only known for some special cases (e.g.,  $k = 2$ ), some allocations have been derived that maximize bounds on the expected opportunity cost of a potentially incorrect selection, when an asymptotically large number of samples is to be allocated.

Corollary 2 of Chick et al. (2001) derives such an one-stage  $\mathcal{L}\mathcal{L}$  allocation, assuming jointly independent, normally distributed outputs with unknown means and known sampling variances that may differ for each system. This policy is analogous to the one-stage  $\mathcal{L}\mathcal{L}$  allocation in Chick and Inoue (2001) that handles the case of unknown means and variances that may differ for each system.

With four adaptations, the one-stage  $\mathcal{L}\mathcal{L}$  allocation can be used as a simulation selection algorithm. One, we note that specifying the prior distributions for the performance of each system obviates the need for the ‘usual’ first stage of sampling that is found in many ranking and selection procedures. Two, for a small to medium number of samples, some of the allocations can be negative. Techniques, such as those used in the  $\mathcal{L}\mathcal{L}$  for the case of unknown sampling variances (Chick and Inoue 2001), can be used to remedy any violations of a non-negativity constraint. Three, the allocation can be made sequential by updating statistics and repeatedly allocating replications until a stopping rule is satisfied. Four, the allocation can be extended to account for discounting by incorporating new stopping rules, such as  $\text{EOC}_1^\gamma$  and  $\text{EOC}_k^\gamma$  in Section 5, that discount the value of information from additional sampling.

These adaptations culminate in the following algorithm.

**Procedure  $\mathcal{L}\mathcal{L}$  (known variances).**

1. Specify prior distributions for the unknown means  $\Theta_i$ , with  $\Theta_i \sim \text{Normal}(\mu_{0i}, \sigma_i^2/t_{0i})$ , for each alternative. Set  $y_{0i} = \mu_{0i}t_{0i}$  for each  $i$ . Include  $\mu_{00} = m$  as an option if a known NPV alternative exists, such as the “do nothing” option with  $m = 0$ . (Set  $\sigma_0^2$  to be very small, e.g.  $10^{-6}$  and  $t_{00}$  to be very large, e.g., 100 years worth of replications, for numerical reasons.)
2. Determine the order statistics, so that  $\mu_{0(0)} \leq \mu_{0(1)} \leq \dots \leq \mu_{0(k)}$ .
3. WHILE stopping rule not satisfied DO:
  - (a) Initialize the set of systems considered for additional replications,  $\mathcal{S} \leftarrow \{0, 1, \dots, k\}$ .
  - (b) For each  $(i)$  in  $\mathcal{S} \setminus \{(k)\}$ : If  $(k) \in \mathcal{S}$  then set  $\lambda_{ik}^{-1} \leftarrow \hat{\sigma}_{(i)}^2/t_{0,(i)} + \hat{\sigma}_{(k)}^2/t_{0,(k)}$ . If  $(k) \notin \mathcal{S}$  then set  $\lambda_{ik} \leftarrow t_{0,(i)}/\hat{\sigma}_{(i)}^2$ .
  - (c) Tentatively allocate a total of  $r$  replications to systems  $(i) \in \mathcal{S}$  (set  $r_{(j)} \leftarrow 0$  for  $(j) \notin \mathcal{S}$ ):

$$r_{(i)} \leftarrow \frac{(r + \sum_{j \in \mathcal{S}} t_j)(\sigma_{(i)}^2 \gamma_{(i)})^{\frac{1}{2}}}{\sum_{j \in \mathcal{S}} (\sigma_j^2 \gamma_j)^{\frac{1}{2}}} - t_{(i)},$$

where

$$\gamma_{(i)} \leftarrow \begin{cases} \lambda_{ik}^{1/2} \phi(d_{ik}^*) & \text{for } (i) \neq (k) \\ \sum_{(j) \in \mathcal{S} \setminus \{(k)\}} \gamma_{(j)} & \text{for } (i) = (k) \end{cases}$$

and  $d_{ik}^* = \lambda_{ik}^{1/2}(\mu_{(k)} - \mu_{(i)})$ .

- (d) If any  $r_i < 0$  then fix the nonnegativity constraint violation: remove  $(i)$  from  $\mathcal{S}$  for each  $(i)$  such that  $r_{(i)} \leq 0$ , and go to Step 3b. Oth-

erwise, round the  $r_i$  so that  $\sum_{i=1}^k r_i = r$  and go to Step 3e.

- (e) Run  $r_i$  additional replications for system  $i$ , for  $i = 1, 2, \dots, k$ . Update the sample statistics,  $t_{0,i} \leftarrow t_{0,i} + r_i$ ;  $y_{0i} \leftarrow y_{0i} +$  sum of  $r_i$  outputs for system  $i$ ;  $\mu_{0i} \leftarrow y_{0i}/t_{0i}$ ; and the order statistics, so that  $\mu_{0(0)} \leq \mu_{0(1)} \leq \dots \leq \mu_{0(k)}$ .
4. Select the system with the best estimated mean.

Depending upon the stopping rule, we refer to Procedure  $\mathcal{L}\mathcal{L}(\text{EOC}_1^\gamma)$  or  $\mathcal{L}\mathcal{L}(\text{EOC}_k^\gamma)$ .

The value of  $r$  in Step 3c is taken to be  $r = 1$  replication per stage for a fully sequential algorithm. The value of  $r$  can be increased if more replications per iteration are desired, e.g., if several replications per stage are run, or if several replications can be run in parallel during each stage. A computational speed-up can be obtained for the allocation, when  $r = 1$ , by ignoring the potential requirement to iterate through Steps 3a-3e, and by directly allocating one replication to the alternative that maximizes  $r_{(i)}$  in the first pass through Step 3c.

Each stopping rule,  $\text{EOC}_1^\gamma$  and  $\text{EOC}_k^\gamma$ , formally identifies the sampling budget  $\beta \geq 1$  that maximizes an approximation to the expected discounted value of continuing to run an additional  $\beta$  replications before selecting a system to implement. The approximations to the expected discounted value require that the  $\beta$  samples be allocated to the  $k$  systems with a one-stage allocation. We do that here by assigning  $r \leftarrow \beta$  and allocating the samples with Steps 3a-3e. The determination of the optimal value of  $\beta$  incurs a computational cost that is associated, for example, with a line-search optimization algorithm for  $\beta$ . A computational speed-up can be obtained by simply checking if there exists a  $\beta \geq 1$  such that the expected discounted value of sampling is positive. If that is the case, then the optimal  $\beta$  certainly has a positive expected discounted value of sampling. In our implementation, we initially solve for the optimal  $\beta$ . If that value exceeds 1, we continue sampling. In the next iteration, we check if a sampling budget of  $\max\{1, \beta - 1\}$  leads to a positive expected discounted value of sampling. If this is so, we continue to sample. If not, we recheck the optimal value of  $\beta \geq 1$  with a line search again.

The left hand sides of the inequalities that determine the stopping rules  $\text{EOC}_1^\gamma$  and  $\text{EOC}_k^\gamma$  are *not* monotonic in  $\beta$ . For example, when comparing  $k = 1$  simulated alternative with a known NPV of 0, if the simulated mean is just below the stopping boundary, the expected reward of a one-step algorithm with  $\beta = 1$  replication might not justify additional sampling. Nevertheless, some values of  $\beta > 1$  may justify additional sampling. It is therefore not optimal to perform a one-step lookahead allocation by only testing if  $\beta = 1$  additional replication is sufficient to justify continuing.

Table 1: The expected discounted reward and average time until selecting a project as a function of the number of independent projects,  $k$ , allocation policy and stopping criterion.

$E[\text{NPV}] \times 10^6$	$k = 3$	4	5	6	7	8	9	10
$\overline{\text{OEDR}}(\vec{\Theta})$	4.44	5.23	5.85	6.35	6.77	7.12	7.42	7.69
$\text{OEDR}(\vec{\Theta})$	4.42	5.20	5.81	6.31	6.72	7.06	7.36	7.62
$\mathcal{L}\mathcal{L}(\text{EOC}_k^\gamma)$	4.43	5.20	5.87	6.39	6.78	7.08	7.41	7.66
$\mathcal{L}\mathcal{L}(\text{EOC}_1^\gamma)$	4.50	5.18	5.78	6.30	6.75	7.09	7.36	7.60
$E[\text{Days}]$								
$\overline{\text{OEDR}}(\vec{\Theta})$	17.4	20.0	22.4	24.5	26.4	28.3	30.0	31.6
$\mathcal{L}\mathcal{L}(\text{EOC}_k^\gamma)$	10.1	8.3	6.6	6.2	6.4	6.3	6.2	6.1
$\mathcal{L}\mathcal{L}(\text{EOC}_1^\gamma)$	10.2	8.2	6.4	6.1	5.9	5.2	5.4	5.4
$\text{PCS}_{i_z}$								
$\overline{\text{OEDR}}(\vec{\Theta})$	0.945	0.935	0.925	0.917	0.909	0.902	0.895	0.889
$\mathcal{L}\mathcal{L}(\text{EOC}_k^\gamma)$	0.967	0.955	0.945	0.938	0.930	0.921	0.916	0.914
$\mathcal{L}\mathcal{L}(\text{EOC}_1^\gamma)$	0.965	0.950	0.943	0.934	0.923	0.905	0.904	0.889

In the numerical experiments of Section 7, we implement the above algorithm with  $r = 1$  replication allocated per stage, and with the preceding computational speedups.

## 7 NUMERICAL RESULTS

### 7.1 Should I Build a Simulation Platform?

A manager can choose to implement one of  $k$  systems directly, or can first choose to build a simulation platform that, once built, would be able to simulate any of the  $k$  alternatives.

Suppose that each of the  $k$  projects has an i.i.d. prior distribution for the unknown mean:  $\text{Normal}(\mu_0, \sigma_i^2/t_0)$  for all  $i$ . We assume that the simulation output for each project is normally distributed with known variance  $\sigma_i = 10^6$ , a cpu time of  $\eta = 20$  min/replication, an annual discount rate of 10%, and no marginal cost for simulations:  $c_i = 0$ .

The top rows of Table 1 display the values of  $\overline{\text{OEDR}}(\vec{\Theta})$  and  $\text{OEDR}(\vec{\Theta})$  as functions of the number of alternatives, when  $\mu_0 = 0$  and  $t_0 = 4$ . These values of  $\overline{\text{OEDR}}(\vec{\Theta})$  and  $\text{OEDR}(\vec{\Theta})$  can be compared with the time and cost required to develop a simulation platform, to decide if a platform warrants building or not, as in Section 4. The data show that the bounds are relatively close for this range of  $k$ .

### 7.2 Simulation Platform Built: How Long to Simulate?

Suppose now that the simulation platform has been built, but that the problem is otherwise the same as in Section 7.1. Table 1 also shows the expected NPV of using Procedure  $\mathcal{L}\mathcal{L}(\text{EOC}_1^\gamma)$  or Procedure  $\mathcal{L}\mathcal{L}(\text{EOC}_k^\gamma)$  to identify the best alternative. Each  $\mathcal{L}\mathcal{L}(\text{EOC}_1^\gamma)$  or  $\mathcal{L}\mathcal{L}(\text{EOC}_k^\gamma)$  cell is based on 6000 i.i.d. problem instances in which a set of unknown means is sampled from their  $\text{Normal}(\mu_0, \sigma_i^2/t_0)$  prior distributions (except for  $k = 1$ , which is based upon

$10^5$  samples, and where the simulation results match the PDE solution with  $E[\text{NPV}] = B(\mu_0, t_0) = 1.99 \times 10^6$ ). For Table 1, each procedure was modified slightly to stop after a maximum of 75 days of sampling, or if the stopping rule is satisfied, whichever comes first.

The top portion of Table 1 shows that  $\mathcal{L}\mathcal{L}(\text{EOC}_k^\gamma)$  and  $\mathcal{L}\mathcal{L}(\text{EOC}_1^\gamma)$  provide estimates of the expected NPVs that are in the range from  $\text{OEDR}(\vec{\Theta})$  to  $\overline{\text{OEDR}}(\vec{\Theta})$ , or within two standard errors of that range. There is a slight advantage for  $\mathcal{L}\mathcal{L}(\text{EOC}_k^\gamma)$  over  $\mathcal{L}\mathcal{L}(\text{EOC}_1^\gamma)$ , as expected, since it samples somewhat more.

The middle portion of Table 1 shows that, on average, both of the sequential  $\mathcal{L}\mathcal{L}$  procedures require much less time than that required by the optimal one-stage procedure that maximizes  $\text{OEDR}(\vec{\Theta})$ . Procedure  $\mathcal{L}\mathcal{L}(\text{EOC}_k^\gamma)$  tends to sample more than  $\mathcal{L}\mathcal{L}(\text{EOC}_1^\gamma)$ , as expected by the construction of the stopping rules. There is no corresponding time duration for  $\overline{\text{OEDR}}(\vec{\Theta})$ , since that figure assumes perfect information instantaneously at no cost.

The bottom portion of Table 1 shows the frequentist probability of correct selection for these procedures, estimated by the fraction of times the ‘true’ best alternative was selected by the procedure. With respect to this criterion,  $\mathcal{L}\mathcal{L}(\text{EOC}_k^\gamma)$  again beats  $\mathcal{L}\mathcal{L}(\text{EOC}_1^\gamma)$ , which in turn beats the optimal one-stage allocation.

For the range of  $k$  tested, more systems means more opportunity to obtain a good system, which means better expected performance. We did not study combinatorially large  $k$  here. We also have not yet studied the use of common random numbers as a variance reduction tool in this context.

## 8 DISCUSSION

Several other results have been obtained. They include:

- the ability to compare the expected NPV of flexible stopping rules, such as those in Section 5, with the expected NPV of rigid stopping rules that are sometimes seen in practice and that specify a fixed simulation analysis deadline;
- additional numerical examples; and
- an improved ‘quick and dirty’ numerical approximation for the optimal simulation stopping time for the simulation selection problem when  $k = 1$  (which improves upon a numerical approximation of Brezzi and Lai (2002) for a related “Bayesian bandit” problem).

## ACKNOWLEDGMENTS

The research of Noah Gans was supported by the Fishman-Davidson Center for Service and Operations Management and the Wharton-INSEAD Alliance.

## REFERENCES

- Brezzi, M., and T. L. Lai. 2002. Optimal learning and experimentation in bandit problems. *J. Economic Dynamics & Control* 27:87–108.
- Chick, S. E., and N. Gans. 2006. Simulation selection problems: Overview of an economic analysis. In *Proc. 2006 Winter Simulation Conference*, ed. L. Perrone, F. Wieland, J. Liu, B. Lawson, D. Nicol, and R. Fujimoto, 279–286. Piscataway, NJ: IEEE, Inc.
- Chick, S. E., and N. Gans. 2008. Economic analysis of simulation selection problems. INSEAD/Wharton Alliance Working Paper, Fontainebleau, France.
- Chick, S. E., M. Hashimoto, and K. Inoue. 2001. Bayesian sampling allocations for selecting the best population with different sampling costs and known variances. In *System and Bayesian Reliability*, ed. M. Xie, T. Z. Irony, and Y. Hayakawa, 333–349. World Scientific.
- Chick, S. E., and K. Inoue. 2001. New two-stage and sequential procedures for selecting the best simulated system. *Operations Research* 49 (5): 732–743.
- de Groot, M. H. 1970. *Optimal statistical decisions*. New York: McGraw-Hill.
- Fu, M., F. W. Glover, and J. April. 2005. Simulation optimization: A review, new developments, and applications. In *Proc. 2005 Winter Simulation Conference*, ed. M. Kuhl, N. Steiger, F. Armstrong, and J. Joines, 83–95. Piscataway, NJ: IEEE, Inc.
- Glazebrook, K. D. 1979. Stoppable families of alternative bandit processes. *J. Appl. Prob.* 16:843–854.
- Gupta, S. S., and K. J. Miescke. 1996. Bayesian look ahead one-stage sampling allocations for selecting the best population. *Journal of Statistical Planning and Inference* 54:229–244.
- Kim, S.-H., and B. L. Nelson. 2006. Selecting the best system. In *Handbook in Operations Research and Management Science: Simulation*, ed. S. Henderson and B. Nelson, Chapter 17, 501534. Elsevier.
- Law, A. M., and W. D. Kelton. 2000. *Simulation modeling & analysis*. 3rd ed. New York: McGraw-Hill, Inc.

## AUTHOR BIOGRAPHIES

**STEPHEN E. CHICK** is a Professor of Technology and Operations Management at INSEAD. 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/>](http://faculty.insead.edu/chick/).

**NOAH GANS** is an Associate 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>](mailto:gans@wharton.upenn.edu).