MANAGEMENT SCIENCE
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0025-1909 | EISSN 1526-5501 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Economic Analysis of Simulation Selection Problems

Stephen E. Chick
INSEAD; Technology & Operations Management Area; Blvd. de Constance; 77305 Fontainebleau France; stephen.chick@insead.edu

Noah Gans
OPIM Department – Wharton; Univ. of Pennsylvania; 3730 Walnut Street, Suite 500; Philadelphia, PA 19104 U.S.A.; gans@wharton.upenn.edu

Ranking and selection procedures are standard methods for selecting the best of a finite number of simulated design alternatives, based on a desired level of statistical evidence for correct selection. But the link between statistical significance and financial significance is indirect and there has been little or no research into it. This paper presents a new approach to the simulation selection problem, one that maximizes the expected net present value (NPV) of decisions made when using stochastic simulation. We provide a framework for answering these managerial questions: When does a proposed system design, whose performance is unknown, merit the time and money needed to develop a simulation to infer its performance? For how long should the simulation analysis continue before a design is approved or rejected? We frame the simulation selection problem as a "stoppable" version of a Bayesian bandit problem that treats the ability to simulate as a real option prior to project implementation. For a single proposed system, we solve a free boundary problem for a heat equation that approximates the solution to a dynamic program that finds optimal simulation project stopping times and that answers the managerial questions. For multiple proposed systems, we extend previous Bayesian selection procedures to account for discounting and simulation-tool development costs.

Key words: simulation, ranking and selection, economics of simulation, optimal stopping, free boundary problem
History: 20 Dec 2007; Revised: 28 May 2008; September 12, 2008

Managers must decide the operating characteristics of their companies' manufacturing, supply chain, or service delivery systems. Often the decision reflects the choice of one among a number of competing designs. To aid their decision-making, managers may use stochastic or discrete event simulation. For a fixed, finite set of alternative designs, one must decide how long to simulate each alternative and, given the simulation results, which design to implement.

A common approach for selecting the best of a finite set of simulated systems uses ranking and selection procedures, which seek to provide a desired level of statistical evidence that the system with the best performance is ultimately selected. A typical measure of statistical evidence is the probability of correct selection (PCS). Good ranking and selection procedures attempt to minimize the mean number of replications that are needed to reach a desired level of statistical evidence for correct selection. This is a flexible approach that allows one to assess a wide variety of operational and other measures of system performance.

But statistical significance is not the same as financial significance, and when system performance and simulation results are themselves financial measures, the maximization of expected net present value (NPV) may be a more appropriate objective (Brealey and Myers 2001). That is, if a manager's goal is to maximize the expected NPV of high-level system design choices, then she is faced with two countervailing costs. On the one hand, uncertainty about the expected NPV of each alternative compels her to simulate more to reduce the opportunity cost associated with an incorrect selection. On the other hand, a simulation analysis itself may incur direct costs, and simulation-driven delays in project implementation may reduce the NPV of the system that is ultimately implemented, due to discounting. Further, standard practice for sound simulation studies (e.g., Law and Kelton 2000, §1.7) does not provide formal guidance via economic principles about whether or not an alternative should be simulated at all.

In this paper, we formulate and solve a simulation selection problem in which the manager seeks to maximize the expected NPV of the system eventually selected, less discounting and analysis costs. Our formulation of the problem 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 (rather than the actual, but unknown, NPV). Section 1 defines the problem and identifies our assumptions, and §2 compares our formulation with more traditional approaches found in the simulation literature.

Section 3 shows that, among procedures that sequentially select systems to simulate and then stop to implement a system, there exists a deterministic, stationary policy that is optimal. Section 4 then provides asymptotic approximations for the optimal expected discounted NPV of the simulation selection problem when there is exactly one simulated alternative. The analysis indicates how long one must simulate before choosing to implement or reject the alternative, given simulation output that is normally distributed with a known variance. The asymptotic regime is reasonable given typical discount rates and simulation run times.

The approximation is determined by the solution of a free boundary problem for a heat equation that shares characteristics with financial and real options. That theory is applied to illustrative simulation selection scenarios in §5 to demonstrate the economic value of our approach, and to show how a manager can use our results to decide whether or not a design proposal warrants the time and money that is required to develop simulation tools.

Section 6 extends the scope of our analysis to problems with more than one simulated alternative. It begins by noting that well-known sufficient conditions for the existence of an optimal "allocation" index, which could simplify the characterization of the optimal simulation selection policy, do not appear to hold. The characterization of an optimal selection procedure policy for multiple systems therefore remains an open question. Nevertheless, §6 extends previous work for Bayesian selection procedures that account for the expected value of information, but not for discounting, to the present context with discounting. In numerical examples, the new policies are shown to be close to optimal.

In summary, this paper presents a new approach to the simulation selection problem, one that maximizes the expected NPV of decisions made when using stochastic simulation. The framework is designed to help answer these managerial questions: When does a proposed system design, whose performance is unknown, merit the time and money needed to develop a simulation to infer its performance? For how long should the simulation analysis continue before a design is approved or rejected? The contributions include: the framing of the simulation selection problem as a "stoppable" version of a Bayesian bandit problem, one that treats the ability to simulate as a real option prior to project implementation; the solution to a free boundary problem for a heat equation that approximates the solution to a dynamic program that finds optimal simulation project stopping times; and the extension of previous Bayesian selection procedures to account for discounting and simulation tool development costs.

The Appendices in the Online Companion provide mathematical proofs and specify the numerical methods used in the paper. They also describe how to handle autocorrelated output from steady-state simulations that are amenable to analysis with batch means, as well as simulation run times that differ for each system. In addition, they present a framework for future work by linking the simulation selection problem to variations of the well-known bandit problem, and improve some numerical approximations for "Bayesian bandits".

1. Simulation Selection Problem Description

A manager seeks to develop one of k projects, labeled $i = 1, 2, \ldots, k$. 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.

1.1. 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, $i^* = \arg\max_i \{\mathrm{E}[X_i]\}$.

We note that, although we model NPVs as simple random variables, the systems that generate them may be quite complex. For example, a particular project's sequence of cash flows may involve the composition of several interrelated random processes describing the evolution of investments, $\mathcal{I}(v)$, revenues, $\mathcal{R}(v)$, and operating costs, $\mathcal{O}(v)$, over time, v. Nevertheless, given a continuous-time discount rate $\delta > 0$, each realization of these processes, ω_i, yields a sample $X(\omega_i) = \int_{v=0}^{\infty} [\mathcal{R}(v, \omega_i) - \mathcal{O}(v, \omega_i) - \mathcal{I}(v, \omega_i)] e^{-\delta v} dv$. Here, v is the time elapsed from the moment a system is selected. (The letter t is used differently below.)

Fox and Glynn (1989) and Appendix D.1 suggest techniques for sampling the $X(\omega_i)$ if the time horizon is truly infinite. Projects that are to be used for only a finite time (e.g., a 5-year usable time horizon) can be implemented with terminating simulations, which effectively set $\mathcal{R}(v, \omega_i) - \mathcal{O}(v, \omega_i) - \mathcal{I}(v, \omega_i)$ to 0 during all but a finite interval. Fixed or random duration delays from the time a project is selected to the time of implementation (due to the need for project approval or startup delays) can similarly be implemented by setting $\mathcal{R}(v, \omega_i)$, $\mathcal{O}(v, \omega_i)$ or $\mathcal{I}(v, \omega_i)$ to 0 during an initial interval.

This approach to modeling delays is valid whenever they are statistically independent of the duration of the simulation analysis that led to the selection. This precludes fixed, pre-scheduled implementation dates, which can occur in practice. Nevertheless, the analysis below suggests that such pre-scheduled implementation dates may themselves be suboptimal, given that they ignore a manager's option to implement earlier or to pursue additional analysis, depending on simulation results obtained up to that fixed date.

In this paper, we assume that the distributions of the X_i's are not known with certainty by the manager. Rather, she believes that a given X_i comes from one of a family of probability distributions, $P_{X_i|\theta_i}$, indexed by a parameter θ_i from a parameter space Ω_{Θ_i}. We model her belief with a probability distribution on θ_i, which we call P_{Θ_i}. For example, the manager may believe that X_i is normally distributed with a known variance, σ_i^2, but unknown mean. Then P_{Θ_i} represents a probability distribution for the mean. To ease notation, we sometimes refer to the distribution as Θ_i. The expected NPV of project $i > 0$ is then $\mathrm{E}[X_i] = \mathrm{E}[X(\Theta_i)] \triangleq \iint X(\theta_i) dP_{X_i|\theta_i} dP_{\Theta_i}$. We denote the vector of distributions for the projects by $\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_k)$.

1.2. Using 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.

We model the running of simulations as occurring at a sequence of discrete stages $t = 0, 1, 2, \ldots$. Let $\mathbf{X}_t$ be the set of all outputs seen through stage t. We represent Bayesian updating of prior beliefs and sample outcomes through time, $\{(\boldsymbol{\Theta}_t, \mathbf{X}_t) \,|\, t = 0, 1, \ldots\}$ 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 Bayes' rule determines the posterior distribution $\Theta_{i,t+1}$, which is a function of the parameter θ_i:

$$dP_{\Theta_{i,t+1}}(\theta_i \,|\, x_{i,t}, \Theta_{i,t}) \; = \; \frac{dP_{X_i\,|\,\theta_i}(x_{i,t} \,|\, \theta_i)\, dP_{\Theta_{i,t}}(\theta_i)}{\int_{\theta_i} dP_{X_i\,|\,\theta_i}(x_{i,t} \,|\, \theta_i)\, dP_{\Theta_{i,t}}(\theta_i)} \qquad \forall\, \theta_i \,\in\, \Omega_{\Theta_i}, \tag{1}$$

while $\Theta_{j,t+1} = \Theta_{j,t}$, and $X_{j,t}$ need not be defined for all $j \neq i$. Thus, the evolution of the manager's beliefs regarding the distribution of outcomes of each project, $\Theta_{i,t}$, 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 θ_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 the simulation runs do take time and do cost money. We assume that the marginal cost of each run of system i is $\$c_i$ and takes η_i units of time to complete. Thus, given a continuous-time discount rate of $\delta > 0$, the decision to simulate system i once costs the manager c_i plus a reduction of $\Delta_i = e^{-\delta\eta_i}$ times the expected NPV of the (unknown) project that is eventually chosen. Note that $\Delta_i < 1$.

There may also be associated up-front costs associated with the development of the simulation tool, itself. It may require time and money to develop an underlying simulation platform, independent of which projects end up being evaluated. Additional costs may be required to be able to simulate particular projects.

This paper initially makes two assumptions regarding the costs of simulation that simplify the analysis. First, we assume that the up-front costs and delays to develop the simulation tools are sunk for all k projects. This is an implicit assumption of all other research on selection procedures. 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 4.4 relaxes the first assumption. Appendix D.2 relaxes the second.

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 the 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 the expected NPV.

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

Then a *selection policy* is the choice of a sequence of simulation runs, a stopping time, and a final project. We define Π to be the set of all *non-anticipating* selection policies, whose choice at time $t = 0, 1, \ldots$ depends only on the history up to t: $\{\boldsymbol{\Theta}_0, \mathbf{X}_0, \boldsymbol{\Theta}_1, \mathbf{X}_1, \ldots, \boldsymbol{\Theta}_{t-1}, \mathbf{X}_{t-1}, \boldsymbol{\Theta}_t\}$. Given prior distributions $\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_k)$ and policy $\pi \in \Pi$, the expected discounted value of the future stream of rewards is

$$V^\pi(\boldsymbol{\Theta}) \;=\; \mathrm{E}_\pi \left[\sum_{t=0}^{T-1} -\Delta^t c_{i(t)} \;+\; \Delta^T X_{I(T),T} \;|\, \boldsymbol{\Theta}_0 = \boldsymbol{\Theta} \right]. \tag{2}$$

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

2. Literature Review

Two broad classes of research are related to this paper. One is the ranking and selection literature, the other is the bandit and optimal stopping literature. Both have substreams.

Branke et al. (2007) review several statistical approaches to ranking and selection. See also Nelson and Goldsman (2001) and Butler et al. (2001). To date, none of the approaches explicitly accounts for discounting costs due to delays in implementation as a result of simulation times, and very few papers explicitly account for the sampling cost of replications.

Chick and Inoue (2001) provide two-stage procedures whose second stage allocation trades off the cost of sampling with an approximation to the Bayesian expected value of information (EVI) of those samples. Sampling costs may differ for each system as may the unknown sample variances. Their work builds upon earlier results of Gupta and Miescke (1996), who examined the case of known sampling variances, and a fixed number of samples to be allocated. The EVI is measured with respect to one of two loss functions, the posterior probability of incorrect selection (PICS), or the posterior expected opportunity cost (EOC) of a potentially incorrect selection. The EOC is a first step for modeling financial value in selection procedures. Branke et al. (2007) show that specific Bayesian procedures that allocate samples with an EOC criterion, and new adaptive stopping rules, perform very effectively for several classes of selection problems.

The indifference-zone (IZ) approach provides a frequentist guarantee of selection procedure effectiveness (Kim and Nelson 2006b). Almost all IZ procedures focus on guarantees for PCS for each problem instance within a given class, and most IZ work ignores sampling costs. An exception is Hong and Nelson (2005), who account for the cost of switching from one system to another and a common sampling cost for each system. In separate work that is related to this paper, Kim and Nelson (2006a) use diffusion approximations for sequential IZ screening procedures to reduce the simulation time required to guarantee a desired PCS.

In another stream of literature, Gittins (1979) offers an early account of optimal dynamic allocation indices (later called Gittins indices) for infinite-horizon, discounted multi-armed bandit problems. Glazebrook (1979) provides sufficient conditions under which these index results apply to reward streams derived from stoppable arms. Gittins (1989) shows that Glazebrook's results hold under a slightly weaker set of assumptions.

In general, Gittins indices are difficult to compute exactly. Chang and Lai (1987) derive approximations for the Gittins index for the infinite-horizon discounted "Bayesian bandit" problem. Brezzi and Lai (2002) use a diffusion approximation for the Gittins index of a Bayesian bandit that is motivated by work of Chernoff (1961) on composite hypothesis testing (also see Breakwell and Chernoff 1964, Chernoff 1965).

This paper uses Chernoff-like diffusion approximations to solve the simulation selection problem (with $k = 1$ system) in an asymptotically optimal way. That asymptotically-optimal solution is shown to provide an improved approximation to the Gittins index of Brezzi and Lai's Bayesian bandit problem (see Appendix E).

We will show that the simulation selection problem, with $k \geq 1$ systems, is an example of what Glazebrook called a *stoppable family of alternative bandit processes*. We indicate that Glazebrook's sufficient conditions for a Gittins index to exist for "stoppable bandits" do not appear to be satisfied, so the existence of an Gittins index for the simulation selection problem is an open question. Still, we show that EOC-based sampling allocations that are analogous to those in Chick and Inoue (2001), together with new stopping rules, are effective solutions for the simulation selection problem.

3. General Results

This section shows that, given mild technical conditions, a simple class of stationary and deterministic policies, which we call "reasonable," is optimal for the simulation selection problem. It then further characterizes the simulation selection problem with $k = 1$ system, to prepare for our approximation results in §4.

We begin by noting that a policy is *stationary* if the action it prescribes, given state $\mathbf{\Theta}_t = (\Theta_{1,t}, \Theta_{2,t}, \ldots, \Theta_{k,t})$, is independent of the time index, t. A policy is *deterministic* if the action it prescribes is never randomized. Blackwell (1965) has shown that, in infinite-horizon problems with discounted rewards, the following conditions ensure that there exists a deterministic, stationary policy that is optimal: 1) given any state and action, expected one-period rewards are finite; 2) the same, finite set of actions is available in all states.

The proof of Lemma 1 shows how the original problem formulation can be transformed to meet these conditions. The potentially finite stopping time of the simulation selection problem is converted to an infinite horizon by converting the one-time reward $\mathrm{E}[X(\Theta_{I(T),T})]$ at the simulation stopping time T into a perpetuity

$(1-\Delta)\mathrm{E}[X(\Theta_{I(T),T})]$ that is received at each period $t \geq T$. Without loss of generality, then, we can restrict our attention to the class of stationary, deterministic selection policies for the infinite-horizon problem.

LEMMA 1. *Suppose there exists a finite* Υ *such that* $\max\{\mathrm{E}[|X(\theta_i)|], |c_i|\} \leq \Upsilon$ *for all* $\theta_i \in \Omega_{\Theta_i}$ *and* $i = 1, 2, \ldots, k$ *for the simulation selection problem in (2). Then there exists a deterministic, stationary policy* $\pi^* \in \Pi$ *that is optimal.*

See Appendix C in the Online Companion for proofs of all claims.

Now consider the simulation selection problem with a single project, i, and no ability to select the "do nothing" option ($i = 0$). For this problem we call the stopping time T_i, let $i(t) = i$ for $t < T_i$ and let $I(t) = i$ for $t \geq T_i$. Thus, π_i determines a stopping time, T_i, and an associated expected value,

$$V_i^{\pi_i}(\Theta_i) \;=\; \mathrm{E}_{\pi_i}\left[\sum_{t=0}^{T_i-1} -\Delta^t c_i \;+\; \Delta^{T_i} X(\Theta_{i,T_i}) \,|\, \Theta_{i,0} = \Theta_i\right]. \tag{3}$$

We denote the optimal stopping policy and stopping time for project i as π_i^* and T_i^*. If expected one-period rewards are uniformly bounded, then there exists a stationary, deterministic policy that is optimal. Further, the optimal value function satisfies the so-called Bellman equation (Bertsekas and Shreve 1996, Prop. 9.8):

$$\begin{aligned} V_i^{\pi_i^*}(\Theta_{i,t}) &= \max\left\{-c_i + \Delta\mathrm{E}[V_i^{\pi_i}(\Theta_{i,t+1}) \,|\, \Theta_{i,t}, t \neq T_i],\; (1-\Delta)\mathrm{E}[X(\Theta_{i,t})] + \Delta V_i^{\pi_i^*}(\Theta_{i,t})\right\} \\ &= \max\left\{-c_i + \Delta\mathrm{E}[V_i^{\pi_i}(\Theta_{i,t+1}) \,|\, \Theta_{i,t}, t \neq T_i],\; \mathrm{E}[X(\Theta_{i,t})]\right\}. \end{aligned} \tag{4}$$

We call $V_i^{\pi_i^*}(\Theta_{i,t})$ the optimal expected discounted reward (OEDR) for the option to simulate alternative i before deciding whether or not to implement it. In (4), the expression $\Theta_{i,t}$ refers to a specific (conditional) distribution that describes the uncertainty about the parameter θ_i, given the first t observations, and $\Theta_{i,t+1}$ is the (random) conditional distribution for θ_i that will be realized after the next sample to be observed, $X_{i,t+1}$, is actually observed. So the expectation $\mathrm{E}[V_i^{\pi_i}(\Theta_{i,t+1}) \,|\, \Theta_{i,t}, t \neq T_i]$ averages over values of $X_{i,t+1}$.

Since setting $T_i = \infty$ is a feasible (though not necessarily optimal) stationary policy, we know that $V_i^{\pi_i}(\Theta_{i,t}) \geq -c_i/(1-\Delta)$. In turn, from (4) it follows that an optimal policy will never choose the right maximand, and stop simulating, if $(1-\Delta)\mathrm{E}[X(\Theta_{i,t})] < -c_i$.

More generally, we call any stationary, deterministic stopping policy, π_i, *reasonable* if $T_i = t < \infty$ implies $(1-\Delta)\mathrm{E}[X(\Theta_{i,t})] \geq -c_i$. Thus a reasonable policy never stops when the one-period expected revenue from a project falls below the cost of sampling. Similarly, a reasonable policy for the entire simulation selection problem has $T = t$ and $I(t) = i$ only if $(1-\Delta)\mathrm{E}[X(\Theta_{I(t),t})] \geq -c_{I(t)}$.

LEMMA 2. *An optimal deterministic, stationary policy for the simulation selection problem is reasonable, almost surely.*

Thus, without loss of generality, we can restrict our attention to the analysis of reasonable policies.

4. A Value Function Approximation for One Alternative

A normative solution to the analysis of a single simulated alternative requires the evaluation of the OEDR, $V_i^{\pi_i^*}(\Theta_{i,t})$. In this section we develop diffusion approximations that provide structural insight into the form of the OEDR and allow for its efficient computation. Our approach follows in the spirit of Chernoff (1961).

This section assumes $k = 1$, so to simplify notation we drop the system index, i. It also assumes that the simulation output X_j is i.i.d. $\texttt{Normal}\,(\theta, \sigma^2)$ for replication $j = 1, 2, \ldots$, with a known finite variance σ^2 and unknown mean θ. We suppose that θ has a prior distribution Θ which is $\texttt{Normal}\,(\mu_0, \sigma_0^2)$. While this assumption may not satisfy the uniform boundedness condition in Lemma 1, the analysis below results in a well-defined finite OEDR when σ_0^2 is finite.

The diffusion approximations are asymptotically appropriate when the effective discount rate over the duration of a simulation replication is small, as is usually the case. Repeated sampling leads to realizations of a scaled Brownian motion with drift.

The calculation of the OEDR involves the solution of a so-called free boundary problem for a heat equation that is obtained from the diffusion approximation. The boundary is “free” since it is determined by equating the two maximands in the value function, rather than being a known, pre-specified boundary. A comparison of the maximands in the continuous-time analogue of (4) determines the free boundary between a continuation set (in which it is optimal to continue simulating a project) and a stopping set (in which it is optimal to stop simulating and to implement the project).

We motivate the diffusion approximation, present a standardized free boundary problem for the diffusion, and solve the problem for the special case of $c = 0$. The solution when $c > 0$ is proven to be a function of the solution when $c = 0$. We then derive the solution to the optimal stopping problem when comparing a single ($k=1$) simulated project with a project that has a known, deterministic NPV. This section concludes by showing how the diffusion approximation can be used to determine whether or not a simulation tool for one alternative should be implemented in the first place.

4.1. Diffusion Approximation for the Output of One System

Define $t_0 = \sigma^2/\sigma_0^2$, and redefine $t = t_0 + n$, where n is the number of simulation observations seen so far for the single system in question. Set $y_{t_0} = t_0\mu_0$ and $y_t = y_{t_0} + \sum_{j=1}^{n} x_j$. This transformation conveniently makes the posterior distribution Θ_t of θ given $x_1, x_2, \ldots, x_n$ a $\texttt{Normal}\,(y_t/t, \sigma^2/t)$ distribution and will help to find an optimal stopping time for (3) when there is $k=1$ system. We will represent Θ_t, given $x_1, x_2, \ldots, x_{t-t_0}$, by (y_t, t), which uniquely determines it. When $X_1, X_2, \ldots, X_{t-t_0}$ have not yet been observed, (Y_t, t) can be thought of both as a real-valued stochastic process and as a random distribution.

Proceeding informally at first, suppose that observations are obtained continuously rather than at discrete intervals, so that Y_t is a Brownian motion with unknown drift θ and variance σ^2 per unit time. If we let B^π be the continuous-time analog of V^{π_i} in (3), we can take the supremum over all non-anticipative policies π with a stopping time $T \geq t_0$ to obtain the continuous-time approximation for the OEDR

$$B(y_{t_0}, t_0) = \sup_{T \geq t_0} \mathrm{E}_\pi \left[-\int_{t_0}^{T} c e^{-\delta(\xi - t_0)} d\xi + e^{-\delta(T-t_0)} D(y_T, T) \right], \tag{5}$$

where discounting is applied to sampling costs and to the terminal reward function, $D(y_T, T) = y_T/T$.

Appendix B shows that $B(y_t, t)$ is the solution to the following heat equation in a continuation set $\mathcal{C}$,

$$0 = -c - \delta B + \frac{y}{t} B_y + B_t + \frac{\sigma^2}{2} B_{yy}, \tag{6}$$

where subscripts on B indicate partial derivatives, and $B(y_t, t) = D(y_t, t)$ outside of $\mathcal{C}$. The free boundary, $\partial\mathcal{C}$, of $\mathcal{C}$ is determined by the following conditions,

$$B(y,t) = D(y,t) = y_t/t, \text{ on } \partial\mathcal{C} \tag{7}$$

$$B_y(y,t) = D_y(y,t), \text{ on } \partial\mathcal{C} \text{ (smooth pasting).}$$

The basic problem is to solve for the value function, B, and the free boundary, $\partial\mathcal{C}$, determined by (6)-(7).

4.2. Standardized Free Boundary Problem for Optimal Stopping

The free boundary problem that is determined by (6)-(7) depends on three parameters, c, δ and σ. In the spirit in which problems with normal distributions are analyzed using z-statistics, we can rescale specific instances of the problem to obtain a standardized free boundary problem for optimal stopping. To do this we define a new time scale, $\tau = \gamma t$. Set $\tau_0 = \gamma t_0$. Let $Z_\tau = \alpha Y_t$ be a scaled motion with $z_{\tau_0} = z_0 = \alpha y_{t_0}$.

Let $\mu = \beta\theta$ be a rescaled drift parameter. So $\mathrm{E}[Z_\tau] = \mu\tau = \mathrm{E}[\alpha Y_t] = \alpha\theta t = \frac{\alpha\mu}{\beta\gamma}\tau$. If $\alpha/\beta\gamma = 1$ then the drift of Z_τ is μ. Also, $\mathrm{Var}[Z_\tau] = \alpha^2\mathrm{Var}[Y_t] = \alpha^2\sigma^2 t = \frac{\alpha^2\sigma^2}{\gamma}\tau$, so Z_τ has unit variance per time unit if $\alpha^2\sigma^2 = \gamma$. Those two moment relations constrain the set of suitable choices of γ, α, β. The third constraint, which is needed to identify the three parameters, is chosen after examining whether or not c equals 0.

4.2.1. Discounting Costs Only ($c = 0, \delta > 0$) Suppose the marginal cost of additional replications is essentially nil ($c = 0$), e.g. if analyst and computer time are considered to be sunk costs, but simulation delays discount a project's value ($\delta > 0$). We first find values γ, α and β that standardize the problem, then introduce a change of variables that will simplify its solution. The discounted terminal reward at time t is

$$e^{-\delta(t-t_0)} D(y_t, t) = e^{-(\tau-\tau_0)\delta/\gamma} \frac{\gamma}{\alpha} \frac{z_\tau}{\tau}.$$

We choose the parameters to standardize both the discount factor in the loss function ($\delta/\gamma = 1$) and the diffusion parameters ($\alpha/\beta\gamma = 1$ and $\alpha^2\sigma^2/\gamma = 1$). This parametrization requires

$$\alpha = \delta^{1/2}\sigma^{-1}, \beta = \delta^{-1/2}\sigma^{-1} \text{ and } \gamma = \delta. \tag{8}$$

A second change of variables will simplify the solution of the resulting standardized free boundary problem,

$$s = 1/\tau = 1/(\delta t) \text{ and } w_s = z_\tau/\tau = \beta y_t/t,$$

with $s_0 = 1/\tau_0 = 1/(\delta t_0)$, and $W(s_0) = w_0 = z_{\tau_0}/\tau_0 = \beta y_{t_0}/t_0$. Then W is a Brownian motion in the $-s$ scale going backward from time $s = s_0$ to time $s = 0$, with initial point (s_0, w_0) (Chernoff 1961).

With these changes of variables and with $c = 0$, solving (6)-(7) is equivalent to solving the following free boundary problem in (w, s) coordinates. Problem (9) is standardized in that it models $c = 0$, $\delta = 1$ and $\sigma = 1$:

$$0 = -\frac{B_1}{s^2} - B_{1,s} + \frac{1}{2} B_{1,ww} \tag{9}$$

$$D(w, s) = \max\{0, w\}$$

$$B_1 = D \text{ and } B_{1,w} = D_w, \text{ on the free boundary } \partial\mathcal{C}.$$

We use $D(w,s) = \max\{0,w\}$, rather than $D(w,s) = w$, because boundary conditions at $s = 0$ are needed to specify the problem. A discounted reward of 0 at time $s = 0$ corresponds to simulating forever. The following theorem characterizes the the solution of the standardized problem (9):

THEOREM 1. *The free boundary $\partial\mathcal{C}$ of the continuation set for the standardized problem in (9) is a function $b_1(s) \geq 0$. Suppose that $c = 0$ and $\delta > 0$. The OEDR B of the original free boundary problem in (6)-(7) is $B(y_{t_0}, t_0) = \sigma\sqrt{\delta}B_1(w_0, s_0) \geq \max\{0, y_{t_0}/t_0\}$, where B_1 solves (9), and the continuation set is $\mathcal{C} = \{(w,s) : w < b_1(s)\} = \{(y,t) : y/t < \sigma\sqrt{\delta}b_1(1/\delta t)\}$.*

The proof of Theorem 1 can be found in Appendix C. The appendix also fills in some details of the above development.

Thus, if $y_t/t < \sigma\sqrt{\delta}b_1(1/\delta t)$ then it is optimal to simulate. The output of the simulation can be used with Bayes' rule to obtain a posterior distribution for the unknown mean, which becomes the prior distribution for the next stage. If $y_t/t \geq \sigma\sqrt{\delta}b_1(1/\delta t)$, then discounting costs outweigh the value of gathering additional information from more simulations, and there is a higher value $B(y_t, t) = y_t/t$ to implementing immediately.

Theorem 2 characterizes the asymptotics of the stopping boundary – its proof shows that $b_1(s)$ is also the optimal stopping boundary of a related problem that was considered by Brezzi and Lai (2002).

THEOREM 2. $b_1(s) \doteq s/\sqrt{2}$ *as* $s \to 0$ *and* $b_1(s) \doteq s^{1/2}\left(2\log s - \log\log s - \log 16\pi\right)^{1/2}$ *as* $s \to \infty$.

Appendix E shows how we compute B_1 and b_1 for the numerical examples below. The computations make use of the following lemma, which is also used below. The lemma's lower bound is obtained by examining *one-stage policies* which work as follows: first β replications are observed; then, if the posterior mean exceeds $-c/\delta$ (=0 here), the system is selected; otherwise it is rejected in favor of infinite simulation replications.

LEMMA 3. *Let ϕ and Φ be the pdf and the cdf of a standard normal distribution. Let $\Psi[s] = \int_s^\infty (\xi - s)\phi(\xi)d\xi = \phi(s) - s(1 - \Phi(s))$ be the Newsvendor loss function for a standard normal distribution. Then*

$$B(y_{t_0}, t_0) \geq \underline{B}(y_{t_0}, t_0) \triangleq \sup_{\beta \geq 0} e^{-\delta\beta}\left(-\frac{c}{\delta} + \left(\frac{\sigma^2\beta}{t_0(t_0+\beta)}\right)^{1/2} \Psi\left[-\left(\frac{y_{t_0}}{t_0} + \frac{c}{\delta}\right) / \left(\frac{\sigma^2\beta}{t_0(t_0+\beta)}\right)^{1/2}\right]\right). \tag{10}$$

4.2.2. Both Sampling and Discounting Costs ($c, \delta > 0$) A similar analysis to that in §4.2.1 can be applied when both sampling and discounting costs are positive. That analysis, along with the proof of the following theorem, is found in Appendix C.

THEOREM 3. *Let $b_1(s)$ be the free boundary and $B_1(w, s)$ be the OEDR for the standardized problem (9) for the case $c = 0, \delta > 0$. Suppose that $c > 0$ and $\delta > 0$, and set $\kappa = \delta^{3/2}\sigma c^{-1}$. The OEDR B of the original free boundary problem in (6)-(7) is*

$$B(y_{t_0}, t_0) = \beta^{-1}\left(B_1(w_0 + 1/\kappa, s_0) - \frac{1}{\kappa}\right) = \sigma\sqrt{\delta}B_1\left(\frac{1}{\sigma\sqrt{\delta}}(\frac{y_{t_0}}{t_0} + \frac{c}{\delta}), \frac{1}{\delta t_0}\right) - \frac{c}{\delta}, \tag{11}$$

and the continuation set is $\mathcal{C} = \{(w, s) : w < b_1(s) - 1/\kappa\} = \{(y, t) : y/t < \beta^{-1}b_1(1/\delta t) - c/\delta\}$.

Theorem 3 says that the OEDR when $c > 0$ is a simple transformation of the OEDR when $c = 0$. It further says that the continuation set for y/t is shifted by $-c/\delta$. This implies that only one free boundary problem must be solved to handle any values of $c \geq 0$ and $\delta > 0$.

Note that the formula $b(t) = \beta^{-1}b_1(1/\delta t) - c/\delta$ for the boundary of $\mathcal{C}$ is valid for both $c = 0$ and $c > 0$. We define $b^{-1}(m) = \sup\{t : b(t) \geq m\}$ for all $m > -c/\delta$. Theorem 2 implies that $b(t)$ is monotone decreasing and continuous for sufficiently large and sufficiently small t. Numerical experiments in Appendix E suggest that $b(t)$ is, in fact, monotone decreasing and continuous for all t (and thus invertible for $m > -c/\delta$) .

4.3. Comparing a Single Simulated System to a Known Alternative

The analysis in §4.2 requires that one either simulate or implement a single system. In this case, given a simulated system whose E[NPV] is far below $-c/\delta$ with high probability, it is optimal to simulate forever, rather than to implement. Alternatively, one may wish either to simulate, to stop and implement the simulated system, or to stop and obtain a known deterministic NPV whose value is m. If the known deterministic alternative is to "do nothing"/maintain the status quo, then $m = 0$. An arbitrary $m \neq 0$ allows for comparisons with a known standard (Nelson and Goldsman 2001) or with the "retirement option" often used to characterize multi-armed bandit problems (Whittle 1980). We therefore address the following generalization of (4).

$$\begin{aligned} V^{\pi^*}(m, \Theta_t) &= \max\left\{m, -c + \Delta\mathrm{E}[V^{\pi}(m, \Theta_{t+1}) \,|\, \Theta_t, t \neq T],\ (1-\Delta)\mathrm{E}[X(\Theta_t)] + \Delta V^{\pi^*}(\Theta_t)\right\} \\ &= \max\{m, -c + \Delta\mathrm{E}[V^{\pi}(m, \Theta_{t+1}) \,|\, \Theta_t, t \neq T],\ \mathrm{E}[X(\Theta_t)]\}\,. \end{aligned} \tag{12}$$

Several results follow directly from the structure of (12) and the results of the previous subsections. First, since an optimal policy is reasonable, $V^{\pi^*}(m,\Theta_t) = V^{\pi^*}(\Theta_t)$ for all $m \leq -c/\delta$. We therefore focus on $m > -c/\delta$. Second, we can develop a diffusion approximation $B(m, y_t, t)$ to $V^{\pi^*}(m,\Theta_t)$ for the case of normally distributed outputs with a known variance. By examining one-stage policies that run β replications and then select a reward of $\max\{m, -c/\delta, y_{t_0+\beta}/(t_0+\beta)\}$, we obtain the following analog of Lemma 3:

$$B(m, y_{t_0}, t_0) \geq \sup_{\beta \geq 0} e^{-\delta\beta} \left(m + \left(\frac{\sigma^2 \beta}{t_0(t_0+\beta)} \right)^{1/2} \Psi \left[-\left(\frac{y_{t_0}}{t_0} - m \right) \left(\frac{\sigma^2 \beta}{t_0(t_0+\beta)} \right)^{-1/2} \right] \right). \qquad (13)$$

Third, a better bound than (13) might be found by noting that the diffusion approximation for (12) is the same as in (6) in the continuation set $\mathcal{C}_m$. The boundary conditions change from (7) to

$$B(m,y,t) = D(m,y,t) \triangleq \max\{m, y/t, -c/\delta\}, \text{ and } B_y(m,y,t) = D_y(m,y,t), \text{ on } \partial\mathcal{C}_m. \qquad (14)$$

The continuation set $\mathcal{C}_m$ is indexed by m as it may, in principal, differ from $\mathcal{C}$.

We define the *m-diffusion problem* to be the free boundary problem that is determined by the heat equation in (6) and the free boundary condition in (14). The lower bound for the solution to the m-diffusion problem in Theorem 4 is based upon the following stopping rule: do not continue sampling if one would stop if the mean were $\max\{y/t, m\}$. With this rule, the maximum number of replications that one should be willing to make is $\lceil b^{-1}(m) \rceil - t_0$ before one stops either to implement the simulated alternative or to take m.

THEOREM 4. *For a fixed m, let $B(m,y,t)$ solve the m-diffusion problem given by (6) and (14). Let $B(y,t)$ be the solution to (6)-(7) in (11), with boundary $b(t) = \beta^{-1} b_1(1/\delta t) - c/\delta$ and continuation set $\mathcal{C} = \{(y,t) : y < b(t)\}$, where β, δ, b_1 are as above. Set $\tilde{t}(t) = t b^{-1}(m)/(b^{-1}(m) - t)$ and $\tilde{t}_0 = \tilde{t}(t_0)$.*

If $m \leq -c/\delta$, then $B(m, y_{t_0}, t_0) = B(y_{t_0}, t_0)$. If $m > -c/\delta$, then $b^{-1}(m)$ is finite and

$$B(m, y_{t_0}, t_0) \geq \underline{B}(m, y_{t_0}, t_0) \triangleq \begin{cases} \max\{y_{t_0}/t_0, m\} & \text{if } t_0 \geq b^{-1}(m) \\ m + \beta^{-1} B_1 \left(\beta(\frac{y_{t_0}}{t_0} - m), 1/\delta\tilde{t}_0 \right) & \text{if } t_0 < b^{-1}(m). \end{cases} \qquad (15)$$

The second alternative of (15) depends upon c, as expected, because $\tilde{t}_0$ is a function of $b^{-1}(m)$, and $b(t) = \beta^{-1} b_1(1/\delta t) - c/\delta$ depends upon c. Note that $\underline{B}(m,y,t)$ is an easily computable function of B_1.

An interesting question that we leave for future work is whether or not $B(m,y,t) = \underline{B}(m,y,t)$. This hypothesis was not rejected by our Monte Carlo tests in §5.

4.4. Should I Develop A Simulation Tool for One Alternative?

The analysis of the previous subsections assumes that the cost of the simulation tools is sunk and that the tools are immediately available for use. Now suppose that the simulation tools have not yet been developed, but that the manager has good estimates of the time and cost required to develop them (setting project scope, data collection, programming, validation, etc.), as well as an estimate of the run times of the simulation replications themselves (e.g., from prior experience with similar projects).

In particular, suppose that $u_0 \geq 0$ years and $\$g_0 \geq 0$ are required to develop the underlying simulation platform that enables the $k = 1$ alternative to be simulated. Let δ_{yr} be the annual discount rate. Then the NPV of having the option to simulate or implement the alternative is:

$$\bar{\mathcal{V}}(\Theta_0) = \max \left\{ \mathrm{E}[X(\Theta_0)], -g_0 + e^{-\delta_{yr} u_0} V^{\pi^*}(\Theta_0) \right\}. \tag{16}$$

The first maximand in (16) is the expected reward from implementing the alternative without building the associated simulation tool, and the second term combines the NPV of developing the simulation tool with the discounted value of the OEDR of the simulation selection problem in §3. The value of $V^{\pi^*}(\Theta_0)$, in turn, can be approximated by the diffusion results on the far right side of (11), which are valid for $c \geq 0$, if there is no choice but to simulate or to implement the simulated alternative, or by (15) if there is an option to stop to implement either the simulated alternative or to select a known deterministic NPV of m.

If $\bar{\mathcal{V}}(\Theta_0) < 0$, then one would neither invest in developing the simulation tools, nor implement the alternative under consideration. If $\bar{\mathcal{V}}(\Theta_0) > 0$ and $\bar{\mathcal{V}}(\Theta_0)$ equals the second maximand of (16), then it is economically optimal to implement the simulation tool. If $\bar{\mathcal{V}}(\Theta_0) > 0$ and $\bar{\mathcal{V}}(\Theta_0)$ equals the first maximand of (16), then it is economically optimal to implement the alternative without developing a tool to simulate it.

5. Sample Simulation Selection Problems

This section applies our results to several illustrative examples with one alternative. Example 1 demonstrates that the optimal stopping rule is more complex than existing stopping rules in ranking and selection. Example 2 shows that positive marginal sampling costs imply a finite amount of time that one should be willing to simulate. These two examples assume that the development cost of the simulation tool is sunk.

Two other examples illustrate the economic value of the approach. Example 3 analyzes whether or not it is optimal to invest in simulation tools in the first place. Example 4 demonstrates the economic value of having a flexible stopping time for simulation, as opposed to a rigid deadline for simulation analysis.

Example 1 examines how large a simulation's output mean must be before one stops to implement a system. Assume that a firm uses a discount rate of 10%/year, that replications of a single simulated alternative have standard deviation $\sigma = \$10^7$ and require $\eta = 20$ minutes to run at no marginal cost ($c = 0$), so that the results of §4.2.1 apply. The simulation time makes the discount rate per replication equal to $\delta = 20 \times 0.10/365/24/60$, so $1/\delta = 2.63 \times 10^5$ replications are required to obtain a scaled time of $\tau = 1$.

The solid line in Figure 1 indicates the boundary between stopping and continuation regions. The first replication at which the sample mean crosses above the stopping boundary is the point at which one stops to implement the proposed system. If the sample mean never were to cross above the stopping boundary, then one would simulate forever.

We examine this stopping boundary further before we compare it to the stopping criteria of some other procedures. For example, our results indicate that simulation should stop after $t = 14$ replications if, at that point, the sample mean is $y_t/t \geq \$10^7$ (corresponding to a z-score of $z = \frac{y_t/t}{\sigma/\sqrt{t}} = 3.7$). If the simulated system is implemented (with $y_t/t > 0$), then the posterior probability of incorrect selection (PICS) is the probability that the unknown mean NPV is less than the value of not implementing any system (which has NPV = 0). Recall that the posterior probability for the unknown mean is $\texttt{Normal}\,(y_t/t, \sigma^2/t)$, with density $p_t(\theta) = \frac{\sqrt{t}}{\sqrt{2\pi\sigma^2}} e^{-(\theta - y_t/t)^2 t/2\sigma^2}$. If $z = 3.7$ when $t = 14$, then PICS $= \int_{-\infty}^{0} p_t(\theta)d\theta = \Phi[-z] = 1 \times 10^{-4}$. If the simulated system is selected as best, but the mean turns out to be $\theta < 0$, then the opportunity cost is $0 - \theta$, and the posterior expected opportunity cost of potentially incorrect selection is EOC $= \int_{-\infty}^{0} (0 - \theta) p_t(\theta) d\theta = 69$.

Note that, as one moves along the stopping boundary, the PICS and EOC change. If one stops after 663 replications (9.2 days) with $y_t/t \geq \$10^6$ ($z = 2.57$, then PICS $= 5.0 \times 10^{-3}$ and EOC $= 615.6$. If one stops after and after 1973 replications (274 days) with $y_t/t \geq \$10^5$ ($z = 1.4$) then PICS $= 9.6 \times 10^{-2}$ and EOC $= 2584$. In this example, then, a greater potential upside means that one is willing to "stop simulating and start building" sooner, but a more stringent level of evidence for correct selection is required (a higher z-score, meaning a lower PICS).

These results allow us to compare our stopping policy with highly effective Bayesian procedures that do not account for discounting in Branke et al. (2007). Those earlier procedures specify a given fixed number of replications, or a fixed PICS or EOC threshold, that determines when stopping should occur. But this example shows that the optimal number of replications, PICS and EOC are not constant thresholds; they change as the simulation results accumulate. Thus, these previous approaches are not optimal for E[NPV]. We also note that the optimal stopping boundary to maximize the E[NPV] of a selection differs from the shapes (e.g., triangular) of stopping regions for several frequentist IZ procedures.

Example 2 shows that the inclusion of marginal costs for sampling compels the analysis to end. Suppose that the variable cost per simulation run is \$3/hour (e.g. for computer time), and all other parameters are as in Example 1. The cost per replication $c = \$3 \times 20/60 = \$1$. We presume that the alternative to stop and "do nothing" is available, with $m = 0$, so that the results of §4.2.2 and §4.3 apply.

Figure 2 shows the original stopping boundary from Figure 1 as a line with long dashes. The solid stopping boundary, drawn $c/\delta \approx$ \$263K below it, accounts for sampling costs. (To allow for negative values, the y-axis is not in log-scale.) That the stopping boundary with $c \geq 0$ falls below that for $c = 0$ reflects the fact that one is willing to simulate for a shorter time: if $y_t/t = \$10^6$, then 6.3 days instead of Example 1's 9.2 days; if $y_t/t = \$10^5$, then 44.2 days instead of Example 1's 274 days).

Thus, a larger c means a willingness to run fewer replications. A larger σ pushes the stopping boundary proportionally higher above the base $-c/\delta$. Appendix D.1 further discusses how σ, c and δ interact to determine the continuation region in the context of stationary simulations.

The figure's horizontal dash-dot line corresponds to the "do nothing" option with a deterministic NPV of $m = 0$. It intersects the stopping line at 5120 replications (71.1 days), the most that one would rationally simulate this system if the goal were to maximize E[NPV]. Beyond that number of replications, one would take the zero option if the posterior mean were $y_t/t < 0$, and one would implement if $y_t/t > 0$.

The dotted line in Figure 2 represents the contour $B(y,t) = 0$. Below that line, the OEDR $B(y,t)$ from (11) is negative. One would lose money by being forced to simulate a poor system if the 0 option were not available. As estimated with (15), the OEDR when the 0 option is available exceeds 0 in that region. There is a positive expected value to sample up to the point where one would stop if the mean were 0.

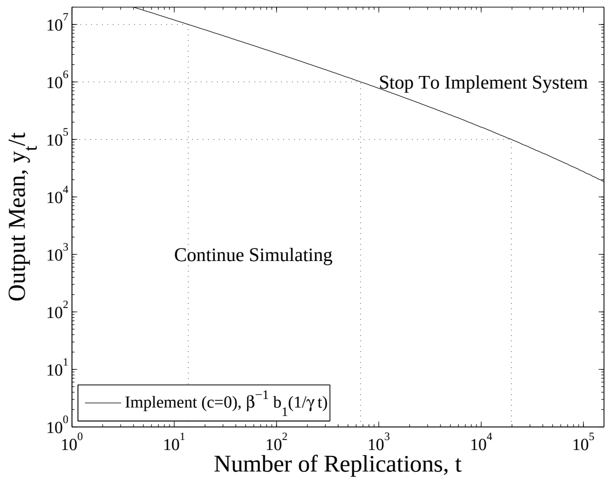


Figure 1 One stops simulating to implement if the sample mean exceeds a stopping boundary ($\sigma = \$10^7$; $\delta = 5.71 \times 10^{-6}$, or 10% per year; $c = 0$).

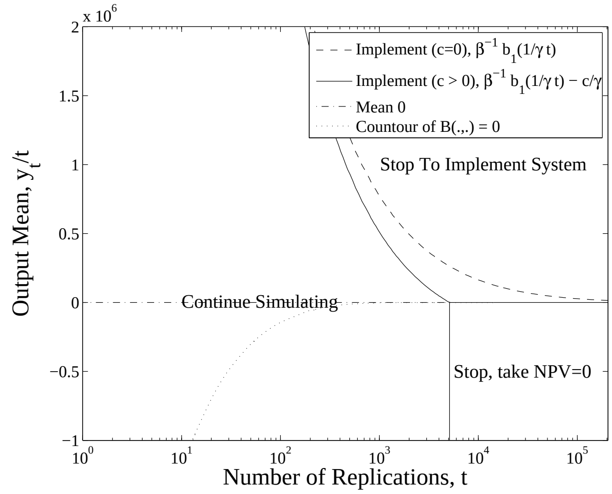


Figure 2 One stops sampling earlier in favor of implementing when the marginal cost of sampling is $\$1$/hour ($\sigma = \$10^7$; $\delta = 5.71 \times 10^{-6}$, or 10% per year).

Example 3 uses §4.4 to assess whether a simulation tool should be developed, assuming that it does not already exist. A manager is considering a system redesign ($k = 1$) as an alternative to continuing with an existing system (the "zero option", which brings no additional revenue beyond the status quo). A validated tool that could simulate the new alternative would require 3 months ($u_0 = 0.25$ years) of time and $g_0 = \$250\text{K}$ to develop. The output of the tool would be the simulated net improvement of the alternative over the mean NPV of continued operation of the current system. The marginal cost of simulation runs is assumed to be negligible ($c = 0$). The firm uses an annual discount rate of $\delta_{yr} = 10\%$. Based upon past simulation experience, a simulation run is predicted to take $\eta = 20$ min, and experience with the existing system leads to an estimate $\sigma = \$10\text{Mil}$ for the standard deviation of the simulated NPV of the alternative.

Should a manager invest time and money in developing the simulation platform? An application of (16) indicates that the answer depends upon the manager's *a priori* assessment of how much better or worse the alternative might be. Suppose that the manager believes that the unknown E[NPV] has, *a priori*, a $\texttt{Normal}\,(\mu_0, \sigma_0^2)$ distribution. For instance, if the manager believes that the alternative has an equal chance of being better or worse, then $\mu_0 = 0$. If the realized difference between the current and the proposed system is of a scale equal to the random noise in the NPV of the existing system, then $\sigma^2 = \sigma_0^2$ and $t_0 = \sigma^2/\sigma_0^2 = 1$. A value of $t_0 = 4$ corresponds to specifying $\sigma_0 = \sigma/2 = \$5\text{Mil}$ in this example.

Figure 3 plots three main policy regions for this example in $\Theta_0 = (t_0, \mu_0)$ coordinates. The boundary of each region depends on the expected value, $-g_0 + e^{-\delta_{yr} u_0} V^{\pi^*}(\Theta_0)$, of building the simulation tool, as well as an application of the simulation selection procedure to learn more before deciding whether or not to implement. The contours represent the absolute value of the difference between $-g_0 + e^{-\delta_{yr} u_0} V^{\pi^*}(\Theta_0)$ and the expected benefit of second policy, which never simulates. The policy to never simulate, and to immediately implement the alternative if and only if the alternative appears favorable, has E[NPV] $= \max(\mu_0, 0)$.

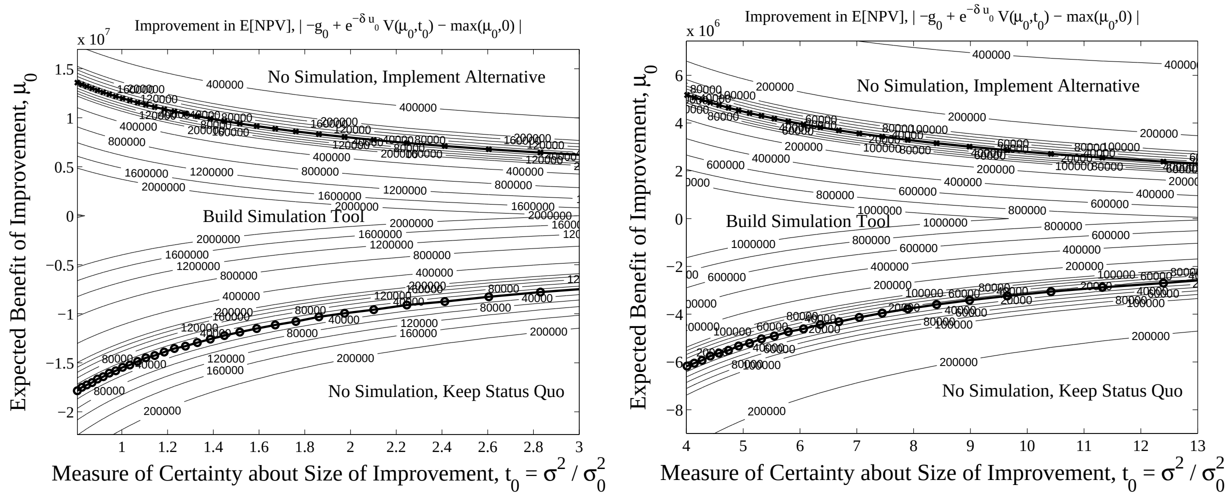
Below the lower bold line with the "○" characters, which is defined by $-g_0 + e^{-\delta_{yr} u_0} V^{\pi^*}(\Theta_0) = 0$, the zero option is more valuable than either maximand in the right hand side of (16). In this region, the manager should therefore not simulate and continue to operate the existing system. The contours in this policy region show the expected loss of simulating, rather than immediately rejecting the alternative.

Above the upper bold line with the "*" characters, the first maximand of (16), which evaluates to μ_0, exceeds both the second maximand and 0. Here, immediately implementing the alternative is preferable to implementing the simulation tool: if $t_0 = 1$, this happens when $\mu_0 >$ \$12Mil; if $t_0 = 4$, this happens when $\mu_0 > \$5.1$Mil. The contours in this upper policy region represent the expected improvement in NPV obtained by immediately implementing the alternative, rather than investing in simulation.

In the middle band of Figure 3, where the alternative is believed to be neither a clear winner nor a serious loser, it is worth the time and investment to develop simulation tools for the analysis. Contours in that band represent the expected benefit of simulating, rather than immediately implementing or rejecting the alternative. For instance, if the manager represents uncertainty about the expected net improvement of the alternative with μ_0=\$4Mil and $t_0 = 4$ (a fair bit of uncertainty), then the improvement in E[NPV] obtained by assessing the alternative optimally with simulation is $-g_0 + e^{-\delta_{yr} u_0} V^{\pi^*}(\mu_0, t_0) - \mu_0$ =\$250K, relative to blindly implementing the alternative. If μ_0=\$0 and $t_0 = 4$, the gain is \$1.7Mil. The more certain the manager is about the mean performance of the alternative (the larger t_0), the narrower the policy region for building the simulation.

Example 4 presumes that the simulation tool from Example 3 has been fully developed. The manager faces a second question – how should the simulation analysis be performed? We consider two ways to perform that analysis. One way is to simulate nonstop until reaching a deadline for a planned decision-making meeting

Figure 3 If the mean performance of the alternative is believed to be too low, one rejects the alternative; if the mean performance is believed to be high, one directly implements. In a middle range, one simulates.



(after, say, a time $t_d = 60$ days), followed by a decision to implement the alternative if the estimated E[NPV] is positive, and to reject the alternative otherwise. Since $c = 0$, the E[NPV] of that plan is

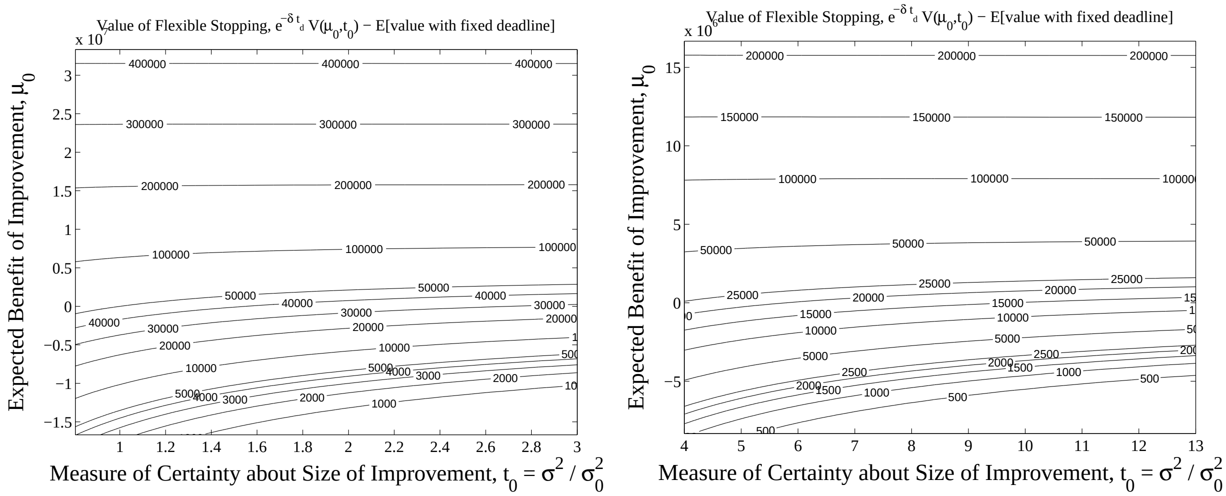
$$\text{E[NPV with fixed deadline]} = e^{-\gamma r}\left(\frac{\sigma^2 r}{t_0(t_0+r)}\right)^{1/2}\Psi\left[-\mu_0/\left(\frac{\sigma^2 r}{t_0(t_0+r)}\right)^{1/2}\right], \tag{17}$$

where $r = t_d \times 24 \times 60/\eta$ is the number of replications that can be run by the deadline (cf. Lemma 3).

Another way to analyze the alternative would be to use the simulation selection procedure developed above: simulate while in the continuation region; if the stopping boundary is reached, then stop to implement the alternative; or reject the alternative if a deterministic fall back of value m were available. Here we set $m = 0$. The alternative is assumed to be implementable two weeks ($t_p = 1/26$ years) after the moment it is selected. This time delay represents coordination time required after the flexible-length analysis, and results in a discounting factor of $e^{-\delta_{yr} t_p}$ times the value of the simulation selection option, $V(\mu_0, t_0)$.

Figure 4 shows the value of a flexible stopping time, relative to that of a rigid decision deadline. For example, if the manager's prior distribution for the unknown E[NPV] has $\mu_0 = \$0$ and $t_0 = 4$, then Figure 4 shows a value of \$25K for flexible stopping out of \$1.7Mil for the value of the simulation analysis option (as at the end of Example 3), for an expected net benefit of 1.5%. When $\mu_0 =$\$4Mil and $t_0 = 4$, that percentage increases to 55K/250K = 22% (cf. Example 3). The value of a simulation analysis with a flexible stopping time, relative to rigid deadlines, increases both with the belief that the alternative is better (larger μ_0), and with uncertainty about the mean NPV (larger $\sigma_0^2 = \sigma^2/t_0$).

Figure 4 The value of flexible stopping for simulation selection, rather than a rigid completion deadline.



6. Multiple Simulated Alternatives

Many simulation studies consider either a small, finite set of distinct systems, or a combinatorially large number of alternatives (e.g., that represent different parameter inputs into a system design structure). This section broadens the scope of our analysis to consider problems with $k > 1$ simulated alternatives.

We begin by recalling the link between simulation selection problems and the multi-armed bandit literature and highlight a difficulty of establishing a so-called Gittins-index result that would greatly simplify the simulation selection problem. Given the lack of such a simplifying result, we then develop bounds for the optimal expected discounted reward of the simulation selection problem. The bounds can be used to determine whether or not simulation tools should be developed. Thus, these results extend the analysis of Example 3 in §5 to problems with $k > 1$ alternatives. We then adapt Chick and Inoue (2001) to develop a sequential sampling algorithm for the simulation selection problem and numerically show that it can deliver near-optimal expected rewards in a timely manner.

6.1. Simulation Selection and the Multi-Armed Bandit Problem

In the multi-armed bandit problem, a decision-maker repeatedly chooses among a fixed set of k independent Markov chains ("arms"), the selected chain alone making a state transition and providing a state-dependent reward. (For a precise definition see Appendix A.) Given certain technical conditions, the problem has an optimal solution with an appealing property: at any time one can calculate a state-dependent index for each chain – called an allocation index or a Gittins index – that is independent of that of the other chains, and it is always optimal to play the chain with the highest index. The simplicity of this type of index rule has attracted significant attention (Gittins 1989) and would be desirable for simulation selection problems.

In fact, the simulation selection problem defined in §1 is similar to that of the multi-armed bandit. Both have discrete-time discounting, independent projects, and Markovian state transitions. At the same time, the simulation selection problem includes a stopping time, T, that is not part of the multi-armed bandit formalism. If, as in the simulation selection problem, a "zero" arm is included, then the multi-armed bandit problem has $k+1$ actions available for all $t = 0, 1, \ldots$. In contrast, for $t \leq T$ the simulation selection problem has $2k+1$ actions available – decide $t < T$ and choose arm $i(t) \in \{1, 2, \ldots, k\}$ to simulate, or decide $t = T$ and choose an arm $I(t) \in \{0, 1, \ldots, k\}$ to implement – and for $t > T$ no actions are available.

The added stopping decision makes the simulation selection problem an example of what Glazebrook (1979) calls a *stoppable family of alternative bandit processes*. The fact that the simulation selection problem is such a "stoppable bandit" problem complicates the question of whether or not an index rule is optimal.

One intuitive solution to the problem with $k > 1$ systems follows a two-step hierarchical structure. First, use the results of §4 to identify an optimal stopping policy for each of the k projects when considered alone. Once these optimal stopping policies are applied, each of k simulated systems behaves as a Markov chain. Then given these k Markov chains, obtained via the fixed application of the k optimal stopping rules, apply the Gittins-index result to sequentially select which system to simulate or implement at a given stage.

Glazebrook (1979, Theorem 3) identified a sufficient condition for such a two-step hierarchical policy to be optimal for stoppable bandit problems. In the context of this paper, that sufficient condition requires that, when each of the k alternatives is in its stopping set when considered individually, the optimal policy for the simulation selection problem with k alternatives, considered together, would also stop.

In Appendix A.3, we construct a simple counter-example to show that Glazebrook's sufficient condition does not hold. In that example, each alternative has statistics that, when considered in isolation, place it barely into the stopping set, so the above two-step hierarchical policy would stop simulating and implement an alternative. Nevertheless, we construct an alternative policy which continues to sample and provides a larger E[NPV]. Thus, the question of whether or not there is an optimal allocation index for the simulation selection problem, and the value of the optimal expected reward, remains open. We therefore take a different tack.

6.2. Should I Develop A Simulation Platform for $k > 1$ Alternatives?

As in §4.4, 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 $k > 1$ 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.

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 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 $\mathbf{r} = (r_1, r_2, \ldots, r_k)$ 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

$$\mu_{00} \triangleq \max\{m, -c_i/\delta : i = 1, 2, \ldots, k\}, \tag{18}$$

and otherwise selects the alternative that maximizes the right hand side of (18). (Recall that $-c_i/\delta$ 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 σ_i^2, but unknown mean whose distribution is $\texttt{Normal}\left(\mu_{0i}, \sigma_i^2/t_{0i}\right)$, with $\mu_{0i} = y_{0i}/t_{0i}$, following the notation in §4. Then the posterior mean that will be realized after the future sampling is done is the random variable (cf. (17))

$$\mathcal{Z}_i \sim \texttt{Normal}\left(\mu_{0i}, \frac{\sigma_i^2 r_i}{t_{0i}(t_{0i}+r_i)}\right). \tag{19}$$

If we consider the allocation to be a function of β and vary β over all possible sampling budgets, we obtain the following lower bound for the OEDR that generalizes Lemma 3 to $k > 1$ projects.

LEMMA 4. *Let $V^{\pi^*}(\mathbf{\Theta})$ maximize (2), and let* $\mathbf{r}$ *be a one-stage allocation. Then*

$$V^{\pi^*}(\mathbf{\Theta}) \geq \underline{OEDR}(\mathbf{\Theta}) \triangleq \sup_{\beta \geq 0} e^{-\gamma\beta} \mathrm{E}\left[\max\{\mu_{00}, \mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_k\}\right] - \sum_{i=1}^{k} r_i c_i. \tag{20}$$

The expectation on the right hand size of (20), in turn, has some easy-to-compute bounds. The bound refers to the order statistics (i) for $i = 0, 1, \ldots, k$ such that $\mu_{0(0)} \leq \mu_{0(1)} \leq \ldots \leq \mu_{0(k)}$.

LEMMA 5. *Let* $\mathbf{r}$ *be a one-stage allocation, let $\Psi[s]$ be as in Lemma 3, let $\sigma^2_{\mathcal{Z},0} = 0$, $\sigma^2_{\mathcal{Z},i} = \frac{\sigma_i^2 r_i}{t_{0i}(t_{0i}+r_i)}$, and $\sigma^2_{\mathcal{Z},i,(k)} = \sigma^2_{\mathcal{Z},i} + \sigma^2_{\mathcal{Z},(k)}$. Then*

$$\mathrm{E}\left[\max\{\mu_{00}, \mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_k\}\right] \geq \mu_{0(k)} + \max_{i: i \neq (k)} \sigma_{\mathcal{Z},i,(k)} \Psi\left[(\mu_{0(k)} - \mu_{0i})/\sigma_{\mathcal{Z},i,(k)}\right] \tag{21}$$

$$\mathrm{E}\left[\max\{\mu_{00}, \mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_k\}\right] \leq \mu_{0(k)} + \sum_{i: i \neq (k)} \sigma_{\mathcal{Z},i,(k)} \Psi\left[(\mu_{0(k)} - \mu_{0i})/\sigma_{\mathcal{Z},i,(k)}\right]. \tag{22}$$

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

$$\overline{\mathrm{OEDR}}(\mathbf{\Theta}) \triangleq \mathrm{E}\left[\max\{\mu_{00}, \mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_k\}\right]. \tag{23}$$

Observe that if $-g_0 + e^{-\gamma u_0}\underline{\mathrm{OEDR}}(\mathbf{\Theta}) > \mu_{0(k)}$, 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). That is because the expected reward from developing the simulation tool and using the allocation $r_i(\beta)$, with the choice of β that determined $\underline{\mathrm{OEDR}}(\mathbf{\Theta})$, would exceed 0.

If $-g_0 + e^{-\gamma u_0}\overline{\mathrm{OEDR}}(\mathbf{\Theta}) < \mu_{0(k)}$, 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).

6.3. Fully Sequential Algorithm with $k > 1$

We now presume that a platform used to simulate the $k > 1$ alternatives has been built, and we seek an effective sequential simulation selection policy. Section 6.1 noted the difficulty of finding the optimal sequential policy. We therefore 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 $\mathcal{LL}$ allocation of Chick and Inoue (2001) that is designed to minimize a bound on the expected opportunity cost (EOC) of a potentially incorrect selection. Gupta and Miescke (1996) showed 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 $\mathcal{LL}$ allocation seeks to maximize the expected undiscounted reward over a finite horizon. Branke et al. (2007) showed that a sequential version of the one-stage $\mathcal{LL}$ allocation is efficient for a variety of selection problems.

Appendix F adapts and extends the $\mathcal{LL}$ 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 $\mathcal{LL}$ allocation of β 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 $\mathcal{LL}$ 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.

The development of §6.2 immediately suggests a mechanism to formalize whether or not there is value to additional sampling. One should continue to sample if $\underline{\text{OEDR}}(\boldsymbol{\Theta}) > \mu_{0(k)}$. This will happen if there is a one-stage allocation of size β that leads to value for continuing to simulate. Unfortunately, the sequential recalculation of $\underline{\text{OEDR}}(\boldsymbol{\Theta})$ that would be required by such a procedure is computationally burdensome.

Fortunately, there is an easy to compute substitute. Substituting the right hand side of (21) for the expectation in the right hand side of (20) leads to an easily computable and analytically justifiable bound.

Stopping rule EOC$_1^\gamma$ (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

$$e^{-\gamma\beta}\left(\mu_{0(k)} + \max_{i:i\neq(k)}\left\{\sigma_{\mathcal{Z},i,(k)}\Psi\left[(\mu_{0(k)}-\mu_{0i})/\sigma_{\mathcal{Z},i,(k)}\right]\right\}\right) - \sum_{i=1}^{k} r_i c_i > \mu_{0(k)}. \tag{24}$$

In practice, EOC$_1^\gamma$ may not be as effective as hoped. It may sample less than is optimal because EOC$_1^\gamma$ 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 (22) into the expectation of (20).

Stopping rule EOC$_k^\gamma$ Continue sampling if and only if there is a budget $\beta \geq 1$ such that

$$e^{-\gamma\beta}\left(\mu_{0(k)} + \sum_{i:i\neq(k)}\sigma_{\mathcal{Z},i,(k)}\Psi\left[(\mu_{0(k)}-\mu_{0i})/\sigma_{\mathcal{Z},i,(k)}\right]\right) - \sum_{i=1}^{k} r_i c_i > \mu_{0(k)}. \tag{25}$$

Appendix F fully specifies how these stopping rules are used with the $\mathcal{LL}$ allocation to solve the simulation selection problem. Depending upon the stopping rule, we refer to Procedure $\mathcal{LL}$(EOC$_1^\gamma$) or $\mathcal{LL}$(EOC$_k^\gamma$). We note that the left-hand sides of (24) and (25) are not monotonic in β, so procedures that use these inequalities must be tested for $\beta \geq 1$, and not for $\beta = 1$ alone.

6.4. Numerical Results

We now extend the numerical examples of §5 by allowing for $k > 1$ alternatives.

Example 5. We extend Examples 1 and 3 by assuming there are $k \geq 1$ projects, each with the same i.i.d. prior distribution for the unknown mean: `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 $\underline{\text{OEDR}}(\mathbf{\Theta})$ and $\overline{\text{OEDR}}(\mathbf{\Theta})$ as functions of the number of alternatives, when $\mu_0 = 0$ and $t_0 = 4$. These values of $\underline{\text{OEDR}}(\mathbf{\Theta})$ and $\overline{\text{OEDR}}(\mathbf{\Theta})$ 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 §6.2. The data show that the bounds are relatively close for this range of k.

Example 6. Suppose now that the simulation platform has been built, but that the problem is otherwise the same as in Example 5. Table 1 also shows the expected NPV of using Procedure $\mathcal{LL}(\text{EOC}_1^\gamma)$ or Procedure $\mathcal{LL}(\text{EOC}_k^\gamma)$ to identify the best alternative. Each $\mathcal{LL}(\text{EOC}_1^\gamma)$ or $\mathcal{LL}(\text{EOC}_k^\gamma)$ cell is based on 6000 i.i.d. problem instances in which a set of unknown means is sampled from their $\texttt{Normal}\,(\mu_0, \sigma_i^2/t_0)$ prior distributions (except for $k = 1$, which is based upon 10^5 samples, and where the simulation results match the PDE solution with E[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 $\mathcal{LL}(\text{EOC}_k^\gamma)$ and $\mathcal{LL}(\text{EOC}_1^\gamma)$ provide estimates of the expected NPVs that are in the range from $\underline{\text{OEDR}}(\boldsymbol{\Theta})$ to $\overline{\text{OEDR}}(\boldsymbol{\Theta})$, or within two standard errors of that range. There is a slight advantage for $\mathcal{LL}(\text{EOC}_k^\gamma)$ over $\mathcal{LL}(\text{EOC}_1^\gamma)$, as expected, since it samples somewhat more.

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

7. Discussion and Conclusions

This paper responds to the question of how to link financial measures, such as a firm's discount rate and the marginal cost of simulations, to the optimal control of simulation experiments that are designed to inform operational decisions. It provides a theoretical foundation, analytical results, and numerical solutions to

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

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.

answer following questions: Should a manager invest time and money to develop simulation tools? For how long should competing systems be simulated before an alternative is selected, or all alternatives are rejected?

This work therefore provides a first link between a managerial perspective on simulation for project selection and the statistical simulation optimization literature. It treats the ability to develop simulation tools, as well as the ability to simulate to gather more information about alternatives, as real options to gather information before committing resources to a design alternative.

The diffusion analysis for a simulation option with $k=1$ alternative assumes normally distributed output with a known variance. For $k \geq 1$ alternatives, we extend earlier ranking and selection work (that minimizes the expected opportunity cost of potentially incorrect selections) to adapt to the current context, that of maximizing the expected discounted NPV of a decision made with simulation. Our results assume that each alternative has a common simulation runtime. Appendix D indicates how these results can be extended to handle autocorrelated output that allows for batch mean analysis, as well as different runtime durations for each system, albeit at a price in terms of optimality. For both of these extensions, further empirical work is needed to assess the quality of the associated approximations, and further theory is needed to assess which assumptions are necessary and sufficient.

The paper also raises several other issues for future study. From a business perspective, we do not address the issue of first-mover advantage or penalties for late implementation of projects due to simulation-analysis induced delays. From a simulation perspective, we do not account for common random numbers (CRN) across systems, a technique that can help sharpen contrasts between systems. Section 3 accounts for unknown variances, but §4 does not. We reserve CRN and unknown variances for future work.

Much current research on simulation optimization focuses on guarantees for the probability of correct selection (PCS) or results for asymptotic convergence. While these are useful properties, this paper suggests that an alternative approach may also be useful: that of maximizing the expected discounted NPV of decisions based on simulation analysis, even at the expense of a potentially incorrect selection. Even with the limitations enumerated above, this new approach to simulation selection accounts for a much fuller accounting of the financial flows that are important to managers.

Acknowledgments

The research of Noah Gans was supported by the Fishman-Davidson Center for Service and Operations Management and the Wharton-INSEAD Alliance. We thank Shane Henderson for feedback on an early draft of this paper, and Ricki Ingalls for discussions about NPV in supply chain simulations. We thank anonymous referees for constructive feedback.

References

Bertsekas, D. P., S. E. Shreve. 1996. *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific.

Blackwell, D. 1965. Discounted dynamic programming. *Annals of Mathematical Statistics* **36** 226–235.

Branke, J., S.E. Chick, C. Schmidt. 2007. Selecting a selection procedure. *Management Science* **53**(12) 1916–1932.

Breakwell, J., H. Chernoff. 1964. Sequential tests for the mean of a normal distribution II (large t). *Ann. Math. Stats.* **35** 162–163.

Brealey, R. A., S. C. Myers. 2001. *Principles of Corporate Finance*. 6th ed. McGraw-Hill.

Brezzi, M., T. L. Lai. 2002. Optimal learning and experimenation in bandit problems. *J. Economic Dynamics & Control* **27** 87–108.

Butler, J., D. J. Morrice, P. W. Mullarkey. 2001. A multiple attribute utility theory approach to ranking and selection. *Management Science* **47**(6) 800–816.

Chang, F., T. L. Lai. 1987. Optimal stopping and dynamic allocation. *Adv. Appl. Prob.* **19** 829–853.

Chernoff, H. 1961. Sequential tests for the mean of a normal distribution. *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. Univ. California Press, 79–91.

Chernoff, H. 1965. Sequential tests for the mean of a normal distribution III (small t). *Ann. Math. Stats.* **36** 28–54.

Chick, S. E., K. Inoue. 2001. New two-stage and sequential procedures for selecting the best simulated system. *Operations Research* **49**(5) 732–743.

Fox, B. L., P. W. Glynn. 1989. Simulating discounted costs. *Management Science* **35**(11) 1297–1315.

Gittins, J. C. 1979. Bandit problems and dynamic allocation indices. *J. Royal Stat. Soc. B* **41** 148–177.

Gittins, J. C. 1989. *Multi-Armed Bandit Allocation Indices*. Wiley, New York.

Glazebrook, K. D. 1979. Stoppable families of alternative bandit processes. *J. Appl. Prob.* **16** 843–854.

Gupta, S. S., 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.

Hong, L. J., B. L. Nelson. 2005. The tradeoff between sampling and switching: New sequential proceduress for indifference-zone selection. *IIE Transactions* **37** 623–634.

Kim, S.-H., B. L. Nelson. 2006a. On the asymptotic validity of fully sequential selection procedures for steady-state simulation. *Operations Research* **54**(3) 475–488.

Kim, S.-H., B. L. Nelson. 2006b. Selecting the best system. S.G. Henderson, B.L. Nelson, eds., *Handbook in Operations Research and Management Science: Simulation*, chap. 17. Elsevier, 501–534.

Law, A. M., W. D. Kelton. 2000. *Simulation Modeling & Analysis*. 3rd ed. McGraw-Hill, Inc, New York.

Nelson, B. L., D. Goldsman. 2001. Comparisons with a standard in simulation experiments. *Management Science* **47**(3) 449–463.

Whittle, P. 1980. Multi-armed bandits and the Gittins index. *J. Royal Stat. Soc. B* **42** 143–149.