Operations Research
Vol. 59, No. 4, July–August 2011, pp. 998–1007
ISSN 0030-364X | EISSN 1526-5463 | 11 | 5904 | 0998
http://dx.doi.org/10.1287/opre.1110.0932

Efficient Nested Simulation for Estimating the Variance of a Conditional Expectation

Yunpeng Sun, Daniel W. Apley, Jeremy Staum
Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208
{yunpengsun2008@u.northwestern.edu, apley@northwestern.edu, j-staum@northwestern.edu}

In a two-level nested simulation, an outer level of simulation samples scenarios, while the inner level uses simulation to estimate a conditional expectation given the scenario. Applications include financial risk management, assessing the effects of simulation input uncertainty, and computing the expected value of gathering more information in decision theory. We show that an ANOVA-like estimator of the variance of the conditional expectation is unbiased under mild conditions, and we discuss the optimal number of inner-level samples to minimize this estimator's variance given a fixed computational budget. We show that as the computational budget increases, the optimal number of inner-level samples remains bounded. This finding contrasts with previous work on two-level simulation problems in which the inner- and outer-level sample sizes must both grow without bound for the estimation error to approach zero. The finding implies that the variance of a conditional expectation can be estimated to arbitrarily high precision by a simulation experiment with a fixed inner-level computational effort per scenario, which we call a one-and-a-half-level simulation. Because the optimal number of inner-level samples is often quite small, a one-and-a-half-level simulation can avoid the heavy computational burden typically associated with two-level simulation.

Subject classifications: nested simulation; analysis of variance; ANOVA; variance components.
Area of review: Simulation.
History: Received August 2009; revisions received March 2010, June 2010, July 2010; accepted August 2010.

1. Introduction

To clarify what we mean by estimating the variance of a conditional expectation, we begin with a mathematical specification and some examples. We consider a random variable X and its conditional distribution given a random vector Z. We are interested in the conditional expectation $M := \mathrm{E}[X \mid Z]$ and its mean $\mu := \mathrm{E}[M] = \mathrm{E}[X]$ and variance $\sigma_M^2 := \mathrm{Var}[M]$. We refer to Z as the *scenario* and to the conditional expectation M as the *value* of the scenario.

For example, consider the problem of assessing the effect of uncertainty about the parameters of distributions used in a simulation model. In particular, suppose the simulation model is of a queueing system, and there is uncertainty about the distributions of service time and of the time between arrivals. A scenario Z consists of the parameters of these distributions. The distribution F_Z of Z represents uncertainty about these parameters. It might be a Bayesian posterior distribution derived from prior beliefs about the system and from observed service times and times between arrivals (Chick 2001, Zouaoui and Wilson 2003). Let X be the time in system of some job, such as the 100th job. (We wish to avoid a discussion of the bias that might arise in studying the steady-state time in system.) The conditional expectation M is the expected time in system of the 100th job given the distribution parameters specified by Z. Its mean μ is the overall expectation of the time in system of the 100th job, taking into account both the stochastic behavior of the system and uncertainty about the parameters of the service time and interarrival time distributions. The variance σ_M^2 of M quantifies uncertainty about the mean time in system of the 100th job due to uncertainty about the parameters (Zouaoui and Wilson 2003). There are also applications in decision theory (Brennan et al. 2007) and financial engineering (Staum 2009) that involve conditional expectation as the value of a scenario.

Next we discuss simulation-based estimation. We assume that we know how to sample from the distribution F_Z of Z and from the conditional distribution $F_{X \mid Z=z}$ of X given $Z = z$ for any z, but that we cannot sample directly from the distribution F_M of M.

If we were interested only in the mean μ, an ordinary, one-level, nonnested simulation would suffice. We could estimate μ by $\sum_{k=1}^{K} X_i/K$, where $X_1, \ldots, X_K$ are sampled independently from the unconditional distribution of X. This can be accomplished as follows: for each $k = 1, \ldots, K$, sample Z_k randomly from F_Z, then sample X_k randomly from $F_{X \mid Z=Z_k}$. This is a one-level simulation in the sense that there is only one realization of X sampled conditional on any particular value of Z; simulating Z is merely an intermediate step in simulating a realization of X. A one-level simulation suffices because the random variables M and X have the same mean, μ. They do not have the same variance, so a one-level simulation of X does not suffice for estimating the variance σ_M^2 of M.

Two-level nested simulation enables estimation of the variance σ_M^2 of F_M, as well as estimation of other quantities, such as probabilities $F_M(y)$ and percentiles $F_M^{-1}(p)$ for $p \in (0, 1)$. Two-level nested simulation works as follows:

- For $k = 1, \ldots, K$:
  —Sample Z_k randomly from F_Z.
  —For $j = 1, \ldots, n_k$:
    Sample X_{kj} randomly from $F_{X \mid Z=Z_k}$.

In the simplest form of two-level nested simulation, each scenario has the same number of inner-level samples: $n_k = n$ for all $k = 1, \ldots, K$. If the inner-level sample size n is sufficiently large, two-level nested simulation provides an accurate estimator $\bar{X}_k := \sum_{j=1}^{n} X_{kj}/n$ of $M_k := \mathrm{E}[X \mid Z = Z_k]$. A straightforward estimator of the variance or 99th percentile of the distribution F_M is the variance or 99th percentile of the empirical distribution $\hat{F}_M$ of $\bar{X}_1, \ldots, \bar{X}_K$, which is given by $\hat{F}_M(y) = \sum_{k=1}^{K} 1\{\bar{X}_k \leqslant y\}/K$. For example, this estimator of the variance is

$$\frac{1}{K}\sum_{k=1}^{K}(\bar{X}_k - \bar{\bar{X}})^2, \tag{1}$$

where $\bar{\bar{X}} := \sum_{k=1}^{K} \bar{X}_k/K$ is an estimator of the mean μ. Such estimators can be badly biased unless the inner-level sample size n is quite large, for the following reason. Define the conditional variance $V := \mathrm{Var}[X \mid Z]$. Then, for all k,

$$\begin{aligned}\mathrm{Var}[\bar{X}_k] &= \mathrm{Var}[\mathrm{E}[\bar{X}_k \mid Z_k]] + \mathrm{E}[\mathrm{Var}[\bar{X}_k \mid Z_k]] \\ &= \sigma_M^2 + \mathrm{E}[V]/n > \sigma_M^2 \end{aligned} \tag{2}$$

if V is nonzero. The estimator (1) exemplifies a typical situation in two-level nested simulation: for its mean-squared error to converge to zero, it is necessary that both the outer- and inner-level sample sizes K and n grow without bound. To make its variance converge to zero, $K \to \infty$ is necessary, and $n \to \infty$ is necessary to make its bias converge to zero. The literature on two-level nested simulation discusses this both in general (Lan et al. 2007) and in detail for estimation of probabilities and quantiles of F_M (Gordy and Juneja 2010, Lee 1998).

Often, the finding is that the inner-level sample size might need to be quite large for a two-level nested simulation estimator to achieve an acceptably low mean-squared error: the average inner-level sample size ranges from several hundred to several thousand in experiments reported by Brennan et al. (2007) and Lan (2010). Gordy and Juneja (2010) reach a different conclusion in studying two-level nested simulation in portfolio risk management: in their examples, the inner-level sample size should be 24 or less. An important message of Gordy and Juneja (2010) is that, contrary to conventional wisdom, the inner-level sample size should be small when simulating a large portfolio. Their finding is specific to portfolio simulation: they show how to make the conditional variance V and thus the bias small when simulating a large portfolio. In contrast, our findings apply to general nested simulation problems and do not require that V be small. *Our main message is that nested simulation supports unbiased estimation of the variance of a conditional expectation, and the optimal inner-level sample size remains bounded as the outer-level sample size grows to make the estimation error go to zero.* We use the term $1\frac{1}{2}$-level simulation to refer to a nested simulation framework in which the inner-level sample size remains constant as the computational budget grows, to distinguish it from a two-level simulation, in which both outer- and inner-level sample sizes grow without bound.

The purpose of this paper is to show how to estimate efficiently the variance of a conditional expectation by nested simulation. First, using an analysis-of-variance (ANOVA) approach, we obtain an unbiased estimator for the variance of the conditional expectation. Zouaoui and Wilson (2003) used a similar approach, but we dispense with their assumption that the conditional variance V is the same for any scenario Z. Our main contribution is to show how to choose the inner-level sample size for maximum computational efficiency. We demonstrate that a $1\frac{1}{2}$-level simulation is optimal, i.e., the optimal inner-level sample size remains bounded as the computational budget grows without bound. We find that this asymptotically optimal inner-level sample size is nearly optimal for many finite budgets encountered in practice, and we discuss how to choose a good inner-level sample size.

The rest of this paper is organized as follows. In §2 we present a general framework and an unbiased estimator. Section 3 is about the optimal inner-level sample size. In §4, we present a method for choosing the inner-level sample size based on a pilot simulation, and we test it numerically. Section 5 contains a more complicated and realistic example to illustrate $1\frac{1}{2}$-level simulation and its benefits. We give conclusions and research directions in §6. Some derivations are deferred to the electronic companion to this paper, which is available as part of the online version at http://or.journal.informs.org/.

2. Derivation of an Unbiased Estimator

The ANOVA framework involves defining new random variables, the *effect* $\tau := M - \mu$ of a scenario and the *error* $\varepsilon := X - M$ associated with observing the effect. Thus, we write the jth inner-level sample conditional on the kth outer-level scenario Z_k as

$$X_{kj} = \mu + \tau_k + \varepsilon_{kj}, \quad \text{where } \tau_k := M_k - \mu \quad \text{and}$$
$$\varepsilon_{kj} := X_{kj} - M_k. \tag{3}$$

The point of this construction is that the effect and error have zero mean, and indeed the error always has zero conditional mean given the scenario. Hence, the error has zero conditional mean given the effect, which makes the effect and error uncorrelated: $\mathrm{E}[\tau\varepsilon] = \mathrm{E}[\tau\,\mathrm{E}[\varepsilon \mid \tau]] = 0 = \mathrm{E}[\tau]\,\mathrm{E}[\varepsilon]$.

The unconditional variance $\mathrm{Var}[X]$ is the sum of two *variance components*, $\sigma_M^2 := \mathrm{Var}[M]$, in which we are primarily interested, and $\sigma_\varepsilon^2 := \mathrm{E}[V]$, the average error variance. Model (3) is a one-way, random effects ANOVA model (Searle et al. 1992), so we use ANOVA methods to estimate σ_M^2. It is not necessary to assume that the effects and errors are independent, which is not generally true in simulation applications. In particular, the conditional error variance $V = \mathrm{Var}[\varepsilon \mid Z]$ is often strongly related to the conditional mean M and thus the effect. For example, in the queuing example mentioned in §1, scenarios that result in larger mean time in system could also have larger variability of the time in system.

ANOVA estimation of variance components refers to the following general strategy:

1. Propose some quadratic forms of the data, often called *sums of squares*.
2. Compute the expectations of the sums of squares as linear functions of the variance components.
3. If the quadratic forms were properly chosen, it is possible to solve the resulting system of linear equations for the variance components as linear functions of the expectations of the sums of squares. Consequently, the corresponding linear functions of the sums of squares are unbiased estimators of the variance components.

The quadratic forms used in standard ANOVA are

$$\mathrm{SS}_\tau = \sum_{k=1}^{K} n_k (\bar{X}_k - \bar{\bar{X}})^2, \quad \text{and}$$

$$\mathrm{SS}_\varepsilon = \sum_{k=1}^{K} \sum_{j=1}^{n_k} (X_{kj} - \bar{X}_k)^2, \tag{4}$$

where

$$\bar{\bar{X}} = \frac{1}{C}\sum_{k=1}^{K} n_k \bar{X}_k, \quad \bar{X}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} X_{kj}, \quad \text{and} \quad C = \sum_{k=1}^{K} n_k.$$

From model (3), we have

$$\bar{X}_k = \frac{1}{n_k}\sum_{j=1}^{n_k} (\mu + \tau_k + \varepsilon_{kj}) = \mu + \tau_k + \bar{\varepsilon}_k, \quad \text{and}$$

$$\bar{\bar{X}} = \frac{1}{C}\sum_{k=1}^{K} n_k(\mu + \tau_k + \bar{\varepsilon}_k), \tag{5}$$

where $\bar{\varepsilon}_k = \sum_{j=1}^{n_k} \varepsilon_{kj}/n_k$. Substituting Equation (5) into SS_τ, while using the facts that the effects and errors all have zero mean and are uncorrelated, and that $\mathrm{E}[\bar{\varepsilon}_k^2] = \sigma_\varepsilon^2/n_k$, we have

$$\begin{aligned}
\mathrm{E}[\mathrm{SS}_\tau] &= \sum_{k=1}^{K} n_k \,\mathrm{E}\left[\left(\tau_k - \frac{1}{C}\sum_{i=1}^{K} n_i \tau_i\right) + \left(\bar{\varepsilon}_k - \frac{1}{C}\sum_{i=1}^{K} n_i \bar{\varepsilon}_i\right)\right]^2 \\
&= \sum_{k=1}^{K} n_k \left\{\left[\left(1 - \frac{n_k}{C}\right)^2 + \frac{1}{C^2}\sum_{i=1,\, i\neq k}^{K} n_i^2\right]\sigma_M^2 \right. \\
&\qquad \left. + \left[\left(1 - \frac{n_k}{C}\right)^2 \frac{1}{n_k} + \frac{1}{C^2}\sum_{i=1,\, i\neq k}^{K} \frac{n_i^2}{n_i}\right]\sigma_\varepsilon^2\right\} \\
&= \left(C - \sum_{i=1}^{K} n_i^2/C\right)\sigma_M^2 + (K-1)\sigma_\varepsilon^2.
\end{aligned} \tag{6}$$

Likewise, substituting $X_{kj} - \bar{X}_k = \varepsilon_{kj} - \bar{\varepsilon}_k$ into SS_ε yields

$$\begin{aligned}
\mathrm{E}[\mathrm{SS}_\varepsilon] &= \sum_{k=1}^{K}\sum_{j=1}^{n_k} \mathrm{E}[(\varepsilon_{kj} - \bar{\varepsilon}_k)^2] \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n_k} \mathrm{E}\left[\left[\left(1 - \frac{1}{n_k}\right)\varepsilon_{kj} - \frac{1}{n_k}\sum_{i=1, i\neq j}^{n_k} \varepsilon_{ki}\right]^2\right] \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n_k} \left[\left(1 - \frac{1}{n_k}\right)^2 + \frac{1}{n_k^2}(n_k - 1)\right]\sigma_\varepsilon^2 \\
&= (C-K)\sigma_\varepsilon^2.
\end{aligned} \tag{7}$$

Solving Equations (6) and (7) for the variance components σ_M^2 and σ_ε^2, and substituting SS_τ and SS_ε for their expectations, yields the unbiased ANOVA estimators

$$\hat{\sigma}_\varepsilon^2 = \frac{\mathrm{SS}_\varepsilon}{C-K} \quad \text{and} \quad \hat{\sigma}_M^2 = \frac{\mathrm{SS}_\tau - (K-1)\hat{\sigma}_\varepsilon^2}{C - \sum_{i=1}^{K} n_i^2/C}. \tag{8}$$

The unbiasedness of these estimators is shown by Searle et al. (1992, p. 71), whose proof does not require that effects and errors are independent and which is valid under the weaker assumption that they are uncorrelated. However, the variance of $\hat{\sigma}_M^2$ is affected by the dependence between the effects and observation errors, as we will see in the next section. For this reason, the standard ANOVA model typically assumes independence between the effects and observation errors, to facilitate testing of hypotheses related to variance components.

In the special case where the inner-level sample size $n_k = n$ is the same for each scenario k, which makes $C = Kn$,

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{K(n-1)}\sum_{k=1}^{K}\sum_{j=1}^{n}(X_{kj} - \bar{X}_k)^2, \quad \text{and}$$

$$\hat{\sigma}_M^2 = \frac{1}{K-1}\sum_{k=1}^{K}(\bar{X}_k - \bar{\bar{X}})^2 - \frac{1}{n}\hat{\sigma}_\varepsilon^2. \tag{9}$$

Zouaoui and Wilson (2003, Equations (20)–(21)) used the estimators in Equation (9) for nested simulation.

3. The Optimal Number of Inner Level Replicates

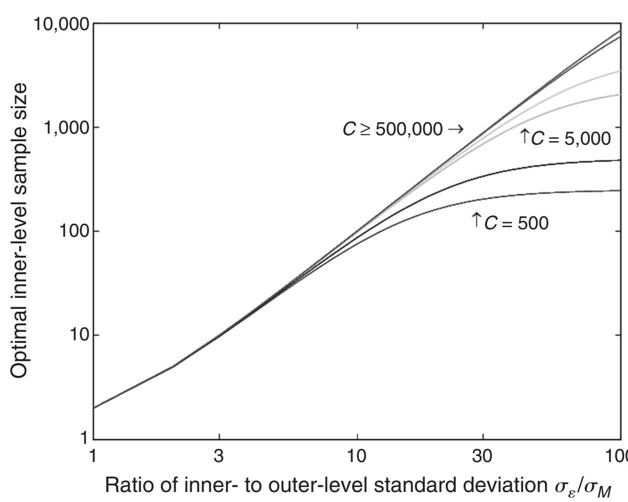
In this section, we study the variance of the estimator $\hat{\sigma}_M^2$ for the purpose of deciding how to choose the number K of outer-level scenarios given a fixed computational budget C. One might take the computational cost to be $K\gamma + \sum_{k=1}^{K} n_k$, where γ is the relative computational expense of generating an outer-level scenario Z compared to generating an inner-level sample X conditional on Z. To simplify the analysis, and because γ is negligible in many simulation applications, we take $\gamma = 0$. Additional analysis, not included in this paper, suggested that all our major conclusions hold when $\gamma > 0$, but with the optimal inner-level sample sizes somewhat larger. For simplicity, we also focus on the special case where the inner-level sample size $n_k = n$ is the same for each scenario k, which makes the computational budget constraint $Kn = C$. Then choosing the number K of scenarios is equivalent to choosing the inner-level sample size n, and the electronic companion shows that the variance of $\hat{\sigma}_M^2(n, K)$ is

$$\begin{aligned}\mathrm{Var}[\hat{\sigma}_M^2(n,K)] = {}& \frac{1}{K}\mathrm{E}[\tau^4] - \frac{(K-3)}{K(K-1)}\sigma_M^4 + \frac{2}{K^2n^2(K-1)}\sigma_\varepsilon^4 \\ &+ \frac{2(K+1)}{K^2n(K-1)}\sigma_M^2\sigma_\varepsilon^2 + \frac{2}{K^2n^3}\mathrm{E}[\varepsilon^4] \\ &+ \frac{2(n^2+(Kn-4)n+3)}{K^2n^3(n-1)}\mathrm{E}[V^2] \\ &+ \frac{4K+2}{K^2n}\mathrm{E}[\tau^2\varepsilon^2] + \frac{4}{K^2n^2}\mathrm{E}[\tau\varepsilon^3], \qquad (10)\end{aligned}$$

where we have added the argument (n, K) to $\hat{\sigma}_M^2$ to make explicit its dependence on the number of inner and outer replicates.

Recall that $\varepsilon = X - M$ is the error associated with one observation of the effect $\tau = M - \mu$, and that ε and τ are not necessarily independent, although they are uncorrelated by definition. Hence minimization of Equation (10) over n for a fixed budget $C = nK$ requires that we know (or estimate) several cross-moments of ε and τ. However, the following arguments, in conjunction with the results in Figure 1, suggest that in many problems a nearly optimal choice of n can be found using a simple asymptotic approximation that is valid for large K. This case is of particular interest, because $K \to \infty$ is necessary for $\mathrm{Var}(\hat{\sigma}_M^2(n, K)) \to 0$.

Figure 1. The inner-level sample size that minimizes the estimator variance in Equation (10) for a fixed computational budget C when errors and effects are independent and effects are normally distributed, vs. the ratio of the standard deviations of errors and effects.



Define the normalized variance $h(n, K) = \mathrm{Var}((nK)^{1/2} \times \hat{\sigma}_M^2(n, K)) = (nK)\mathrm{Var}(\hat{\sigma}_M^2(n, K))$. When comparing two different choices of (n, K), each having the same budget $C = nK$, the one with the smaller value of $h(n, K)$ is preferable. If we intend to use large values of K, then we might consider using the value of n that minimizes the asymptotic normalized variance defined as

$$\begin{aligned}h(n) &= \lim_{K\to\infty} h(n,K) \\ &= n(\mathrm{E}[\tau^4] - \sigma_M^4) + \frac{2}{n-1}\mathrm{E}[V^2] + 4\mathrm{E}[\tau^2\varepsilon^2], \qquad (11)\end{aligned}$$

which follows directly from Equation (10). The unique (noninteger) value of $n \geqslant 1$ that minimizes $h(n)$ is

$$\begin{aligned}n^* &= 1 + \sqrt{\frac{2\mathrm{E}[V^2]}{\mathrm{E}[\tau^4] - \sigma_M^4}} = 1 + \sqrt{\frac{2\mathrm{E}[V^2]}{\sigma_M^4(\kappa_M - 1)}} \\ &= 1 + \sqrt{\frac{2(\sigma_\varepsilon^4 + \mathrm{Var}[V])}{\sigma_M^4(\kappa_M - 1)}}, \qquad (12)\end{aligned}$$

where $\kappa_M = \mathrm{E}[\tau^4]/\sigma_M^4$ denotes the kurtosis of the distribution of M. Equation (12) follows by setting $h'(n) = 0$ and by noting that $h''(n) = 4(n-1)^{-3}\mathrm{E}[V^2] > 0$ for all $n > 1$. Speaking loosely, we will refer to n^* in Equation (12) as the asymptotically optimal inner-level sample size; we next make the notion of an asymptotically optimal inner-level sample size more precise.

An inner-level sample size must be an integer that is greater than one, and n^* may not have these properties. Suppose that the kurtosis $\kappa_M > 1$. Because of the convexity of h on $(1, \infty)$,

$$n^{**} := \arg\min_{n=2,3,\ldots} h(n) = \begin{cases} 2 & \text{if } n^* < 2 \\ \lfloor n^* \rfloor & \text{if } n^* \geqslant 2 \text{ and } h(\lfloor n^* \rfloor) \leqslant h(\lceil n^* \rceil) \\ \lceil n^* \rceil & \text{if } n^* \geqslant 2 \text{ and } h(\lfloor n^* \rfloor) \geqslant h(\lceil n^* \rceil) \end{cases}.$$

The following theorem states that, in the limit as the computational budget $C \to \infty$, the policy of setting the inner-level sample size to n^{**} is as good as any policy of setting the inner-level sample size as a function of C. The remarkable finding that the asymptotically optimal inner-level sample size is a finite constant, as opposed to being unbounded as the budget C grows, is the basis for the phrase "$1\frac{1}{2}$-level simulation." The second part of the following theorem states that the variance reduction ratio of $1\frac{1}{2}$-level simulation, compared to a simulation in which the inner-level sample size goes to infinity as the budget increases, itself goes to infinity as the budget increases.

Theorem 1. *If $\kappa_M > 1$, then for any sequences $\{C_i\}_{i\in\mathbb{N}}$, $\{n_i\}_{i\in\mathbb{N}}$, and $\{K_i\}_{i\in\mathbb{N}}$ of natural numbers such that $C_i \to \infty$ as $i \to \infty$ and for all $i \in \mathbb{N}$, $n_i > 1$, $K_i > 2$, and $n_i K_i = C_i$,*

$$\limsup_{i\to\infty} \frac{\mathrm{Var}[\hat{\sigma}_M^2(n^{**}, \lfloor C_i/n^{**}\rfloor)]}{\mathrm{Var}[\hat{\sigma}_M^2(n_i, K_i)]} \leqslant 1.$$

Furthermore, if $n_i \to \infty$ as $i \to \infty$, then

$$\lim_{i\to\infty} \frac{\mathrm{Var}[\hat{\sigma}_M^2(n^{**}, \lfloor C_i/n^{**}\rfloor)]}{\mathrm{Var}[\hat{\sigma}_M^2(n_i, K_i)]} = 0.$$

Proof. From Equation (10), it follows that for any integers $n > 1$ and $K > 2$,

$$\begin{aligned} &nK\mathrm{Var}[\hat{\sigma}_M^2(n, K)] \\ &\quad \geqslant n(\mathrm{E}[\tau^4] - \sigma_M^4) + \frac{2}{n-1}\mathrm{E}[V^2] + \left(4 + \frac{2}{K}\right)\mathrm{E}[\tau^2\varepsilon^2] \\ &\qquad - \frac{8}{Kn(n-1)}\mathrm{E}[V^2] + \frac{4}{nK}\mathrm{E}[\tau\varepsilon^3] \\ &\quad > h(n) - \frac{8}{Kn(n-1)}\mathrm{E}[V^2] + \frac{4}{nK}\mathrm{E}[\tau\varepsilon^3]. \end{aligned}$$

Therefore, for all $i \in \mathbb{N}$, $C_i\mathrm{Var}[\hat{\sigma}_M^2(n_i, K_i)] > h(n_i) - (8/(C_i(n_i-1)))\,\mathrm{E}[V^2] + (4/C_i)\,\mathrm{E}[\tau\varepsilon^3]$, which implies

$$\liminf_{i\to\infty} C_i \mathrm{Var}[\hat{\sigma}_M^2(n_i, K_i)] \geqslant \liminf_{i\to\infty} h(n_i). \tag{13}$$

Given $\kappa_M > 1$, for all $i \in \mathbb{N}$,

$$h(n_i) \geqslant n_i(\mathrm{E}[\tau^4] - \sigma_M^4) \geqslant \sigma_M^4(\kappa_M - 1) > 0, \tag{14}$$

so both sides of Equation (13) are strictly positive. As $i \to \infty$, $C_i\mathrm{Var}[\hat{\sigma}_M^2(n^{**}, \lfloor C_i/n^{**}\rfloor)]$ converges to the limit $h(n^{**})$ given by Equation (11). Therefore,

$$\begin{aligned} &\limsup_{i\to\infty} \frac{\mathrm{Var}[\hat{\sigma}_M^2(n^{**}, \lfloor C_i/n^{**}\rfloor)]}{\mathrm{Var}[\hat{\sigma}_M^2(n_i, K_i)]} \\ &\quad = \limsup_{i\to\infty} \frac{C_i\mathrm{Var}[\hat{\sigma}_M^2(n^{**}, \lfloor C_i/n^{**}\rfloor)]}{C_i\mathrm{Var}[\hat{\sigma}_M^2(n_i, K_i)]} \\ &\quad = \frac{\lim_{i\to\infty} C_i\mathrm{Var}[\hat{\sigma}_M^2(n^{**}, \lfloor C_i/n^{**}\rfloor)]}{\liminf_{i\to\infty} C_i\mathrm{Var}[\hat{\sigma}_M^2(n_i, K_i)]} \\ &\quad = \frac{h(n^{**})}{\liminf_{i\to\infty} h(n_i)}. \end{aligned}$$

Because $h(n^{**})$ is the smallest value attained by h at any integer greater than one, this proves the first part of the theorem.

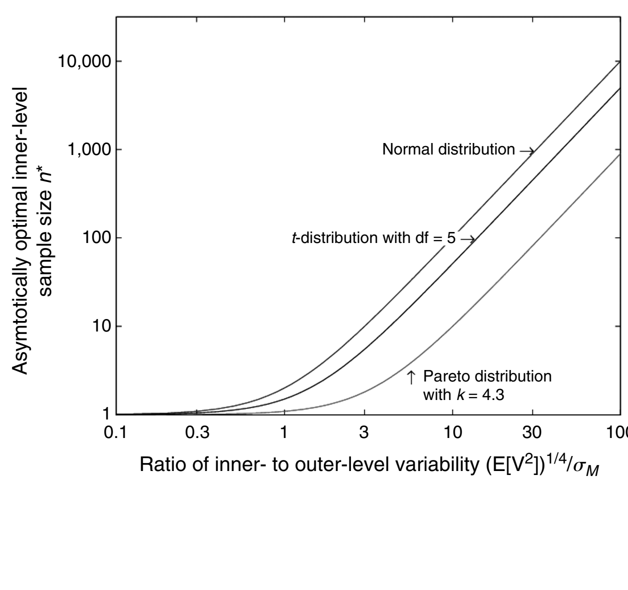
If $n_i \to \infty$ as $i \to \infty$, then Equation (14) implies $\liminf_{i\to\infty} h(n_i) \geqslant \liminf_{i\to\infty} n_i(\mathrm{E}[\tau^4] - \sigma_M^4) = \infty$ and thus

$$\lim_{i\to\infty} \frac{\mathrm{Var}[\hat{\sigma}_M^2(n^{**}, \lfloor C_i/n^{**}\rfloor)]}{\mathrm{Var}[\hat{\sigma}_M^2(n_i, K_i)]} = 0. \quad \square$$

Equation (12) shows that the asymptotically optimal inner-level sample size n^* depends only on the average inner-level variance $\sigma_\varepsilon^2 = \mathrm{E}[V]$, the cross-scenario variability $\mathrm{Var}[V]$ of inner-level variance, the outer-level variance σ_M^2, and the outer-level kurtosis κ_M. A smaller inner-level sample size is better when the inner-level variance is smaller or less variable across scenarios, or when the outer-level distribution has higher variance or kurtosis. The middle expression in Equation (12) shows that n^* is a function of κ_M and the ratio of $\mathrm{E}[V^2] = \mathrm{E}[(\mathrm{Var}[X \mid Z])^2]$ to $\sigma_M^4 = (\mathrm{Var}[\mathrm{E}[X \mid Z]])^2$. Figure 2 shows how n^* depends on the outer-level kurtosis κ_M and on the fourth root of this ratio; if the inner-level variance $V = \mathrm{Var}[X \mid Z]$ does not depend on the scenario, then $\mathrm{Var}[V] = 0$ and the fourth root $(\mathrm{E}[V^2]/\sigma_M^4)^{1/4} = \sigma_\varepsilon/\sigma_M$ is the ratio of inner- to outer-level standard deviations. The kurtoses of the normal, t, and Pareto distributions included in Figure 2 are 3, 9, and 251, respectively—a wide range of values. For $\kappa_M \geqslant 251$ and $\mathrm{E}[V^2] \leqslant \sigma_M^4$, n^* must be rounded up to 2, which is the smallest inner-level sample size that supports unbiased estimation of σ_M^2. For $\kappa_M \geqslant 3$ and $\mathrm{E}[V^2]^{1/4}/\sigma_M \leqslant 3$, $n^* \leqslant 10$; for $\kappa_M \geqslant 3$ and $\mathrm{E}[V^2]^{1/4}/\sigma_M \leqslant 10$, $n^* \leqslant 100$. These inner-level sample sizes are much smaller than what many practitioners typically use.

Because Equation (12) gives an inner-level sample size n^* that is asymptotically optimal as the computational budget C grows, one might wonder how large C must be before n^* is nearly optimal for a finite budget C. Figure 1 answers this question for a special case that fits a standard ANOVA framework: errors and effects are independent, and effects are normally distributed. In this case, $\mathrm{Var}[V] = 0$ and $\kappa_M = 3$, so Equation (12) becomes $n^* = 1 + \sigma_\varepsilon^2/\sigma_M^2$. We obtained the optimal inner-level sample size for a finite

Figure 2. The inner-level sample size n^* that is asymptotically optimal for large computational budgets, given by Equation (12), vs. a ratio of inner- to outer-level variability, for three different outer-level distributions.



budget C by minimizing Equation (10), which can be evaluated explicitly in this special case, in which errors are normally distributed with constant variance $V = \sigma_\varepsilon^2$. The asymptotically optimal inner-level sample size n^* is close to that which is optimal for a finite budget unless the budget is extremely small and the ratio $\mathrm{E}[V^2]^{1/4}/\sigma_M = \sigma_\varepsilon/\sigma_M$ of the inner- to outer-level standard deviations is quite large, compared to typical values we have seen in two-level simulation problems.

In light of this, our recommendation to practitioners who want to estimate the variance of a conditional expectation via two-level simulation is simply to use n^* given by Equation (12) as the inner-level sample size, regardless of the computational budget. This requires estimates of unknown quantities in Equation (12), or guesses for them. The next section presents a method for estimating n^* based on simulation output.

4. Pilot Estimation

Suppose we have the output of a pilot simulation experiment in which there are K_0 scenarios and an inner-level sample size of n_0. To use Equation (12), we plug in estimates of three quantities: $\mathrm{E}[V^2]$, σ_M^4, and $\mathrm{E}[\tau^4]$. The result is that we choose the inner-level sample size

$$\hat{n}^* = \left\lceil 1 + \sqrt{\frac{2\,\widehat{\mathrm{E}[V^2]}}{\widehat{\mathrm{E}[\tau^4]} - \widehat{\sigma_M^4}}} \right\rceil. \tag{15}$$

Here we propose estimators that have some justification, although they are not optimal. We estimate $\mathrm{E}[V^2] = \mathrm{E}[(\mathrm{E}[\varepsilon^2 \mid Z])^2]$ by

$$\widehat{\mathrm{E}[V^2]} = \frac{1}{K_0}\sum_{k=1}^{K_0}\left(\frac{1}{n_0-1}\sum_{j=1}^{n_0} X_{kj}^2 - \frac{(\sum_{j=1}^{n_0} X_{kj})^2}{n_0}\right)^2.$$

A natural estimator of σ_M^4 is

$$\widehat{\sigma_M^4} = (\hat{\sigma}_M^2)^2 = \left(\frac{\mathrm{SS}_\tau}{n_0(K_0-1)} - \frac{\mathrm{SS}_\varepsilon}{n_0K_0(n_0-1)}\right)^2,$$

where the forms of SS_τ and SS_ε are given in Equation (4), but in the present context we substitute n_0 for n and K_0 for K in Equation (4). Finally, we estimate $\mathrm{E}[\tau^4]$ by

$$\widehat{\mathrm{E}[\tau^4]} = \frac{K_0^4}{(K_0-1)^4 + (K_0-1)} \times \left\{\frac{1}{K_0}\sum_{k=1}^{K_0}(\bar{X}_k - \bar{\bar{X}})^4 - \frac{3(K_0-1)(2K_0-3)}{K_0^3}\hat{\sigma}_M^4 - \frac{6(K_0-1)^4 + 6(K_0-1)}{K_0^4 n_0}\widehat{\mathrm{E}[\tau^2\varepsilon^2]}\right\},$$

where $\widehat{\mathrm{E}[\tau^2\varepsilon^2]}$ is an estimate of $\mathrm{E}[\tau^2\varepsilon^2]$. For simplicity, we use an estimator that is natural in the special case where τ and ε are independent:

$$\widehat{\mathrm{E}[\tau^2\varepsilon^2]} = \widehat{\sigma_\varepsilon^2}\sigma_M^2 = \left(\frac{\mathrm{SS}_\varepsilon}{K_0(n_0-1)}\right)\left(\frac{\mathrm{SS}_\tau}{n_0(K_0-1)} - \frac{\mathrm{SS}_\varepsilon}{n_0K_0(n_0-1)}\right).$$

The justification of the estimator $\widehat{\mathrm{E}[\tau^4]}$ is as follows. First, denote

$$\bar{\tau} := \frac{1}{K_0}\sum_{k=1}^{K_0}\tau_k \quad \text{and} \quad \bar{\bar{\varepsilon}} := \frac{1}{K_0}\sum_{k=1}^{K_0}\bar{\varepsilon}_k = \frac{1}{K_0 n}\sum_{k=1}^{K_0}\sum_{j=1}^{n}\varepsilon_{kj},$$

and observe

$$\mathrm{E}\left[\frac{1}{K_0}\sum_{k=1}^{K_0}(\bar{X}_k - \bar{\bar{X}})^4\right] = \frac{1}{K_0}\sum_{k=1}^{K_0}\mathrm{E}(\bar{X}_k - \bar{\bar{X}})^4 = \frac{1}{K_0}\sum_{k=1}^{K_0}\mathrm{E}[(\tau_k - \bar{\tau} + \bar{\varepsilon}_k - \bar{\bar{\varepsilon}})^4].$$

A derivation in the electronic companion, starting with (EC.8), shows that $\mathrm{E}[(\tau_k - \bar{\tau} + \bar{\varepsilon}_k - \bar{\bar{\varepsilon}})^4]$ is given by (EC.9). From this, it follows that

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{K_0}\sum_{k=1}^{K_0}(\bar{X}_k - \bar{\bar{X}})^4\right] \\
&= \frac{(K_0-1)^4 + (K_0-1)}{K_0^4}\mathrm{E}[\tau^4] + \frac{3(K_0-1)(2K_0-3)}{K_0^3}\sigma_M^4 \\
&\quad + \frac{3(K_0-1)(2K_0-3)}{K_0^3 n_0^2}\sigma_\varepsilon^4 + \frac{6(K_0-1)(2K_0-3)}{K_0^3 n_0}\sigma_M^2\sigma_\varepsilon^2 \\
&\quad + \frac{(K_0-1)^4 + (K_0-1)}{K_0^4 n_0^3}\mathrm{E}[\varepsilon^4] \\
&\quad + \frac{3(n_0-1)((K_0-1)^4 + (K_0-1))}{K_0^4 n_0^3}\mathrm{E}[V^2] \\
&\quad + \frac{4(K_0-1)^4 + 4(K_0-1)}{K_0^4 n_0^2}\mathrm{E}[\tau\varepsilon^3] \\
&\quad + \frac{6(K_0-1)^4 + 6(K_0-1)}{K_0^4 n_0}\mathrm{E}[\tau^2\varepsilon^2].
\end{aligned}$$

We approximate

$$\begin{aligned}
&\mathrm{E}\left[\sum_{k=1}^{K_0}(\bar{X}_k - \bar{\bar{X}})^4/K_0\right] \\
&\approx \frac{(K_0-1)^4 + (K_0-1)}{K_0^4}\mathrm{E}[\tau^4] + \frac{3(K_0-1)(2K_0-3)}{K_0^3}\sigma_M^4 \\
&\quad + \frac{6(K_0-1)^4 + 6(K_0-1)}{K_0^4 n_0}\mathrm{E}[\tau^2\varepsilon^2]
\end{aligned} \tag{16}$$

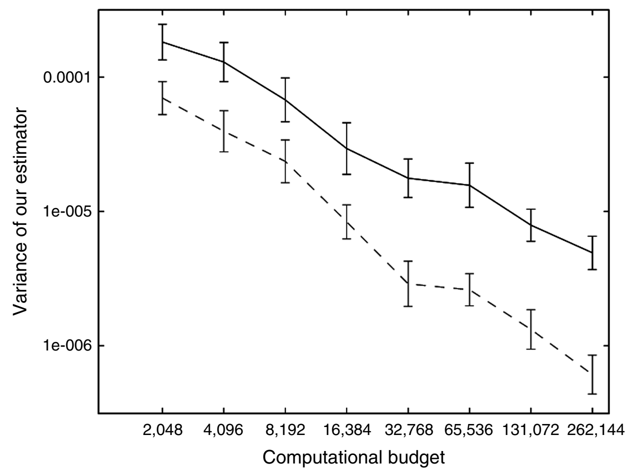
because if n_0 and K_0 are growing large, the terms on the right side of Equation (16) are $\mathcal{O}(1)$, $\mathcal{O}(1/K_0)$, and $\mathcal{O}(1/n_0)$, respectively, and the terms that we dropped all decrease at rates faster than $1/n_0$. Solving Equation (16) for $\mathrm{E}[\tau^4]$ and substituting estimators for expected values leads to the estimator $\widehat{\mathrm{E}[\tau^4]}$ given above. A more careful analysis, leading to better estimators of $\mathrm{E}[V^2]$, σ_M^4, and $\mathrm{E}[\tau^4]$, could lead to better estimation of n^* via pilot simulation, particularly when the pilot simulation has a small budget.

To demonstrate that this pilot estimation method can choose an inner-level sample size that leads to substantial variance reduction compared to an existing two-level simulation method, we ran an experiment using Example 1 of Steckley and Henderson (2003). In this example, the conditional expectation M has a $\beta(4,4)$ distribution, and the conditional distribution of the noise ε given the scenario Z is always normal with mean 0 and variance 0.5. Steckley and Henderson's purpose is estimating the density of a conditional expectation; for more on this topic, see Steckley (2006). We compare the variance of our estimator $\hat{\sigma}_M^2$ when using the inner-level sample sizes chosen by Steckley and Henderson for efficiency in estimating the density of M to the variance of $\hat{\sigma}_M^2$ when using the inner-level sample size $\hat{n}^*$ chosen by pilot estimation for efficiency in estimating σ_M^2. We compare our inner-level sample sizes to theirs because they have a simple, concrete formula for choosing the inner-level sample size as a function of the budget C in this two-level nested simulation problem, even though the formula is not designed for efficient estimation of σ_M^2. Moreover, their formula makes the inner-level sample size proportional to $C^{2/7}$, which is slower growth than the $C^{1/3}$ rate that is MSE-optimal in the cases analyzed by Gordy and Juneja (2010) and Lee (1998). In this sense, the Steckley-Henderson formula is the closest we can find in the simulation literature to $1\frac{1}{2}$-level simulation. The numerical results reported below illustrate that our pilot estimation method works well enough to choose a $1\frac{1}{2}$-level simulation experiment design that is substantially better for our purpose than its closest relative among two-level simulation experiment designs.

Steckley and Henderson let the budget C range from 2,048 to 262,144, and set the inner-level sample size $n_C = \lfloor 30C^{2/7} \rfloor$. We used 10% of the total budget C for a pilot simulation with an arbitrarily chosen small inner-level sample size $n_0 = 8$ and $K_0 = \lfloor 0.1 \times C/n_0 \rfloor$ outer-level scenarios. We computed $\hat{n}^*$ in Equation (15), and then threw out the data from the pilot simulation. We then ran a main simulation with a budget of 90% of C, inner-level sample size $n = \hat{n}^*$, and $K = \lfloor 0.9 \times C/\hat{n}^* \rfloor$ scenarios. We do not necessarily advocate throwing out data from a pilot simulation, but this provides the toughest test for the value of pilot estimation. We measured the estimator variance by running 50 independent macro-replications of the whole experiment. Each macro-replication yielded a single realization of $\hat{\sigma}_M^2(n_C, \lfloor C/n_C \rfloor)$ and a single realization of $\hat{\sigma}_M^2(\hat{n}^*, \lfloor 0.9 \times C/\hat{n}^* \rfloor)$. By computing sample variances over the 50 macro-replications, we obtained estimates of $\mathrm{Var}[\hat{\sigma}_M^2(n_C, \lfloor C/n_C \rfloor)]$ and $\mathrm{Var}[\hat{\sigma}_M^2(\hat{n}^*, \lfloor 0.9 \times C/\hat{n}^* \rfloor)]$.

The results of this experiment appear in Figure 3. The error bars in the figure are computed via the delta method. Even for the smallest budget used by Steckley and Henderson, $C = 2{,}048$, pilot estimation is advantageous in choosing the inner-level sample size. The variance reduction ratio of $\mathrm{Var}[\hat{\sigma}_M^2(n_C, \lfloor C/n_C \rfloor)]$ to $\mathrm{Var}[\hat{\sigma}_M^2(\hat{n}^*, \lfloor 0.9 \times C/\hat{n}^* \rfloor)]$ is about 2.6. While Steckley and Henderson use inner-level sample size $n_{2{,}048} = 264$, the average of $\hat{n}^*$ in these 50 macro-replications was about 8. The values of $\hat{n}^*$ were quite variable across macro-replications with $C = 2{,}048$, ranging from 4 to 49, but there was still a marked variance reduction compared to an inner-level sample size chosen for the purpose of estimating the density of M. As the budget C increases, the variability of $\hat{n}^*$ decreases: for $C \geqslant 65{,}536$, $\hat{n}^*$ was 7 or 8 in each of the macro-replications. As expected in light of Theorem 1, the variance reduction ratio also increases as the budget C increases: for $C = 262{,}144$, the variance reduction ratio is about 8.1.

Figure 3. Comparison of our estimator's variance when using the inner-level sample size of Steckley and Henderson (2003) (solid line), vs. when choosing the inner-level sample size n by our pilot estimation method (dashed line).



Despite the success of pilot estimation in this particular example, in our experience, it is sometimes difficult to estimate n^* accurately with a pilot simulation that is computationally inexpensive compared to the main simulation. In the example treated in the next section, we had to get a good estimate of n^* from the output of a simulation with a large budget and large inner-level sample size. We believe a practitioner who has previously dealt with similar simulation problems by running simulation experiments with a large budget and large inner-level sample size might be able to estimate n^* well from the output of those experiments. Also, by improving upon the estimator $\hat{n}^*$ in Equation (15), it might be possible to estimate n^* using a small pilot simulation. We leave the further analysis and development of methods for choosing n^* to future research.

5. Illustrative Example and Numerical Results

In this section, we provide numerical results demonstrating the increased computational efficiency of $1\frac{1}{2}$-level simulation compared to two-level simulation in an illustrative example drawn from financial engineering. In this example, the goal is to estimate the variance of the profit and loss (P&L) that a trading strategy would produce, by simulating the strategy before actually using it. One point of using such a complicated example is to show that the ANOVA framework, although simple, is flexible enough to accommodate even complicated examples.

The example is of *delta-hedging* a portfolio that is short one European put option and one European call option, with the same strike price Q and maturity T, on an underlying stock, whose price is assumed to follow a geometric Brownian motion. Delta-hedging is a trading strategy in which one adds $-\Delta$ shares of stock to the original portfolio, where Δ is the sensitivity of the original portfolio to the stock price, i.e., the partial derivative of the original portfolio's value with respect to the stock price. The purpose of hedging is to lower the risk of the portfolio by making the new portfolio less sensitive to changes in the stock price than the original portfolio was. Specifically, the variance of the P&L of the new portfolio at a future time T should be less than the variance of the P&L of the original portfolio at T. The hedging strategy consists of self-financing trading in a risk-free money market account with interest rate r and in the underlying stock, at equally spaced times $t_0 = 0, t_1, \ldots, t_{s-1}$, where $t_s = T$ is the options' maturity. The example is very similar to one used by Baysal et al. (2008, §3). Here we focus on formulating the example in a way that fits our ANOVA framework. The scenario Z is a path taken by the stock price over time and the P&L in that scenario is the conditional expectation $M = \mathrm{E}[X \mid Z]$, where the random variable X, given in Equation (18) below, can be interpreted as the P&L that would result in this scenario if one were to hedge using a noisy estimate of Δ.

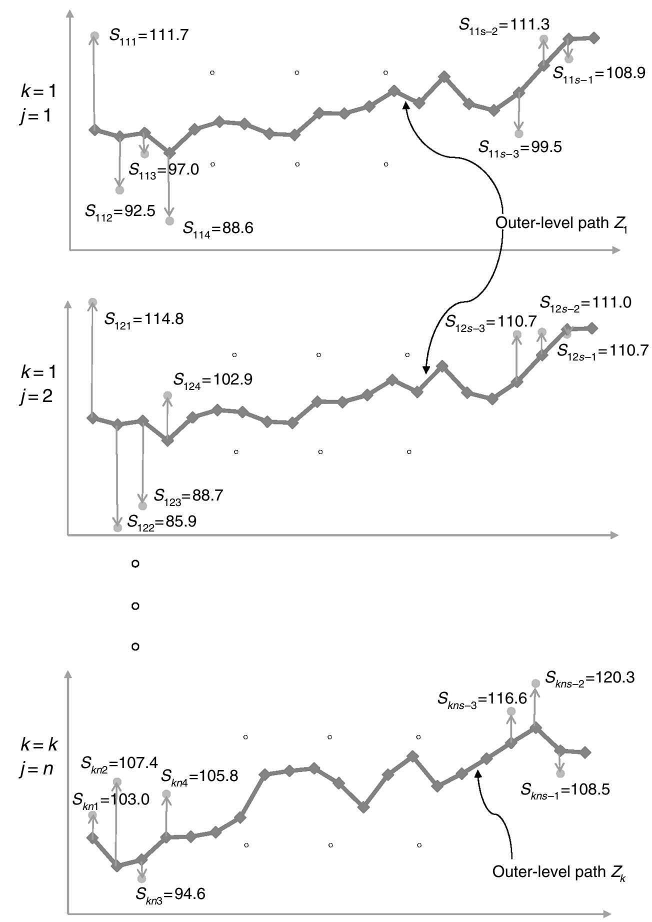
At time t_i, the number of shares of stock in the hedging strategy is updated to $-\Delta_i$, where Δ_i is the sensitivity at time t_i of the original portfolio to the stock price: Δ_i is a function of t_i and S_i. The amount in the money market account is chosen to satisfy the self-financing condition. As shown in Baysal et al. (2008, §3), the P&L of the hedged portfolio at time T is

$$(p_0 + \Delta_0 S_0)e^{rT} + \sum_{i=1}^{s}(\Delta_i - \Delta_{i-1})S_i e^{r(T-t_i)} - |S_s - Q|, \quad (17)$$

where p_0 and $|S_s - Q|$ are, respectively, the initial price and the payoff of the options. Thus, the P&L is a function of the path $S_1, \ldots, S_s$ of stock prices at times $t_1, \ldots, t_s = T$. This path is the scenario Z in our ANOVA framework.

When the stock follows geometric Brownian motion, there is a formula for Δ, which allows the P&L to be computed as an explicit function of the path. Thus, for this particular example, we can compute an accurate estimate of the variance of P&L by one-level simulation. In general, a formula for Δ is not available, and one uses nested simulation to estimate the variance of P&L. The inner level provides estimates of Δ at every time step on every path. The P&L that results from the path $Z = (S_1, \ldots, S_s)$ is the conditional expectation $M = \mathrm{E}[X \mid Z]$ in our ANOVA framework, where the random variable X is given in Equation (18). Figure 4 illustrates the nested simulation. Each outer-level scenario Z_k is a path of stock prices $S_{k1}, \ldots, S_{ks}$. Conditional on this scenario, an inner-level sample X_{kj} involves simulating a collection of stock prices $\{S_{kji}\}_{i=1,\ldots,s-1}$. They do not constitute another path, rather S_{kji} is a stock price at time T simulated conditional on the stock price at time t_i being S_{ki}. The stock prices $\{S_{kji}\}_{i=1,\ldots,s-1}$ are used to provide estimates of Δ at each time t_i on the kth path.

Figure 4. Illustration of nested simulation in the delta-hedging example.



We next exhibit a random variable X such that the P&L in scenario Z is $M = \mathrm{E}[X \mid Z]$. The inner level of simulation is based on pathwise estimation of Δ_i as the sensitivity of the portfolio value to the stock price S_i, which is unbiased under some conditions (Glasserman 2003, §7.2). A pathwise estimator of Δ_i is

$$\psi_i = -e^{-r(T-t_i)} \frac{\tilde{S}}{S_i} \mathrm{sign}(\tilde{S} - Q),$$

where $\tilde{S}$ has the risk-neutral conditional distribution of S_s given S_i. The random variable

$$X = (p_0 + \Delta_0 S_0)e^{rT} + \sum_{i=1}^{s}(\psi_i - \psi_{i-1})S_{ki}e^{r(T-t_i)} - |S_s - Q| \tag{18}$$

can be interpreted as the P&L in the scenario $Z = (S_1, \ldots, S_s)$ if one were to use the hedge ratio ψ_i instead of Δ_i. Its conditional expectation $M = \mathrm{E}[X \mid Z]$ is the P&L given in Equation (17) because of the unbiasedness of the pathwise sensitivity estimation. In the context of the nested simulation illustrated in Figure 4, $Z_k = (S_{k1}, \ldots, S_{ks})$,

$$\psi_{kji} = -e^{-r(T-t_i)}\frac{S_{kji}}{S_{ki}}\operatorname{sign}(S_{kji} - Q), \quad \text{and}$$

$$X_{kj} = (p_0 + \Delta_0 S_0)e^{rT} + \sum_{i=1}^{s}(\psi_{kji} - \psi_{k,j,i-1})S_{ki}e^{r(T-t_i)} - |S_{ks} - Q|,$$

where S_{kji} has the risk-neutral conditional distribution of the stock price at time T given that the stock price at time t_i is S_{ki}. Each panel of Figure 4 shows the simulated stock prices used in one scenario Z_k and one inner-level sample X_{kj} generated conditional on Z_k. The top two panels involve the same scenario Z_1, while the bottom panel involves a different scenario Z_K.

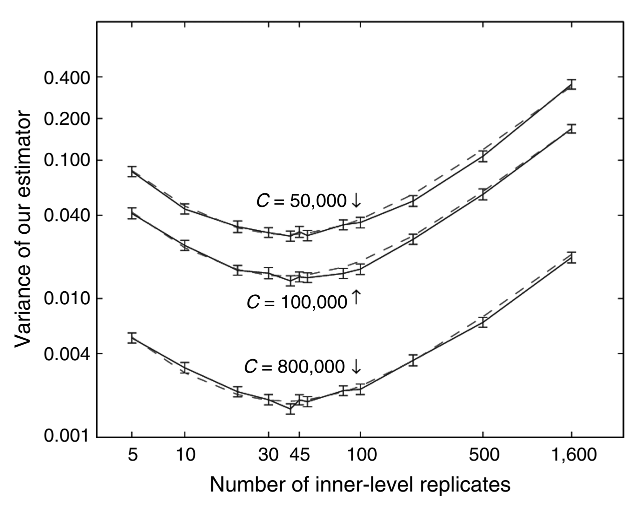
In implementing the example, we have assumed that $\psi_{kj0} = \Delta_0$, the initial delta, is known to high accuracy. Because it is common to all paths, which share the same value of S_0, there is little additional cost in estimating it very accurately. The example would yield similar results if ψ_{kj0} were simulated in the same way as ψ_{kj1}.

Figure 5 illustrates the benefit of $1\frac{1}{2}$-level simulation by showing how the variance of the ANOVA estimator $\hat{\sigma}^2_M$ depends on the inner-level sample size n given a fixed computational budget C. For each pair of n and C, we used 1,000 macro-replications to assess the variance of $\hat{\sigma}^2_M$. Based on the output of a simulation experiment with $K_0 = 100$ outer-level scenarios and inner-level sample size $n_0 = 10{,}000$, using methods described in §4, we estimated the asymptotically optimal inner-level sample size n^* of Equation (12) by $\hat{n}^* = 45$. We then ran nested simulations with different inner-level sample sizes n to see how the variance of $\hat{\sigma}^2_M$ with $n = \hat{n}^*$ compares to the variance with other choices of n. This exercise also demonstrated good agreement of the formula for $\mathrm{Var}[\hat{\sigma}^2_M]$ in Equation (10), where estimates were substituted for unknown quantities, with the direct estimates of $\mathrm{Var}[\hat{\sigma}^2_M]$ based on macro-replications. The numerical results indicate that $\hat{n}^* = 45$ is indeed nearly optimal for the computational budgets considered here. These results provide some validation for our analysis of $\mathrm{Var}[\hat{\sigma}^2_M]$ and n^*. The finding that n^* is near 45 is striking because 45 is a much smaller inner-level sample size than would ordinarily be used in two-level simulation in such an example. In this example, to attain a relative root mean square error of 1% in estimating the P&L M at the inner level would require an inner-level sample size of about 1,600. Figure 5 shows that using $n = 1{,}600$ instead of $n = 45$ makes the variance of $\hat{\sigma}^2_M$ increase dramatically: when the computational budget $C = 800{,}000$, this makes the variance increase by a factor of about 12. Put another way, to attain the same accuracy in estimating σ^2_M by $\hat{\sigma}^2_M$ that is attained with budget $C = 800{,}000$ and $n = 45$, if we were to use $n = 1{,}600$ then we would require a budget of over 10 million.

6. Conclusions and Research Directions

Our principal findings are twofold. First, the ANOVA estimator $\hat{\sigma}^2_M$ of Equation (8) or Equation (9) is an unbiased estimator of the variance of a conditional expectation in nested simulation. Second, this implies that where the inner-level sample size is the same for all scenarios, it is optimal for it to remain bounded as the computational budget grows, leading to the concept of $1\frac{1}{2}$-level simulation. Our recommendation for the nested simulation problems most often encountered in practice is simply to use the asymptotically optimal inner-level sample size n^* given by Equation (12), or its estimator $\hat{n}^*$ given by Equation (15). This sample size n^* is often much smaller than that which would be needed for accurate estimation of the conditional expectation in all scenarios, which is unnecessary for the

Figure 5. Variance of the ANOVA estimator $\hat{\sigma}^2_M$ in the delta-hedging example, as a function of the inner-level sample size given a fixed computational budget, for three different computational budgets C. The solid curves give point estimates of the variance, and the error bars are 95% confidence intervals for the variance. The dashed curves represent the formula for the variance given in Equation (10), with estimates substituted for unknown quantities.



purpose of estimating the variance of the conditional expectation. The smaller sample size can greatly increase computational efficiency.

We believe there is promise in extensions of the central idea presented in this paper to functionals other than variance of the distribution F_M of the conditional expectation M. We showed how to construct an unbiased estimator of $\sigma_M^2 = \text{Var}[M]$ and hence $\text{E}[M^2]$ if the inner-level sample size $n \geqslant 2$. Likewise, it is not hard to show how to construct an unbiased estimator of $\text{E}[M^m]$ if $n \geqslant m$: tools mentioned in Douillet (2009, §5) might be useful in this task. Thus we conjecture that $1\frac{1}{2}$-level simulation would be optimal for estimation of any moment of the conditional expectation M. However, it would be harder to choose the optimal inner-level sample size.

Unbiased estimation of the moments suggests using moment-based approximations of other functionals of F_M, for example, using the Cornish-Fisher expansion to approximate quantiles, or using the Taylor expansion of a function f to approximate $\text{E}[f(M)]$. A different idea is to apply the technique of deconvolution used in signal processing: if effects and errors in model (3) are independent, the distribution of $X = M + \varepsilon$ is the convolution of the distributions of M and of ε. Then F_M can be estimated by estimating F_X and F_ε and "deconvolving" them. This approach might be viable for those simulation problems in which the conditional distribution of error does not vary much across scenarios. It seems that a promising domain for deconvolution would be nested simulation problems with low outer-level variability that are challenging because the inner-level variability is very high compared to the outer-level variability: the right part of Figure 2 shows that these problems call for a large inner-level sample size when estimating σ_M^2. It remains to be seen what advantages these approaches might have over the method, described in the introduction, of running a two-level simulation and estimating a functional of F_M by evaluating that functional on $\hat{F}_M$.

7. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Acknowledgments

This research was partially supported by National Science foundation grants CMMI-0555485 and CMMI-0758557. The authors thank the associate editor and two anonymous referees for their comments, suggestions, and corrections, which enabled the authors to improve the paper substantially. The authors also thank Paul Glasserman and Shane Henderson for suggesting the research direction involving the Taylor expansion.

References

Baysal, R. E., B. L. Nelson, J. Staum. 2008. Response surface methodology for hedging and trading strategies. S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler, eds. *Proc. 2008 Winter Simulation Conf.* IEEE Press, Piscataway, NJ, 629–637.

Brennan, A., S. Kharroubi, A. O'Hagan, J. Chilcott. 2007. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Medical Decision Making* **27**(4) 448–470.

Chick, S. E. 2001. Input distribution selection for simulation experiments: Accounting for input uncertainty. *Oper. Res.* **49**(5) 744–758.

Douillet, P. L. 2009. Sampling distribution of the variance. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, R. G. Ingalls, eds. *Proc. 2009 Winter Simulation Conf.* IEEE Press, Piscataway, NJ, 403–414.

Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

Gordy, M. B., S. Juneja. 2010. Nested simulation in portfolio risk measurement. *Management Sci.* **56**(10) 1833–1848.

Lan, H. 2010. Two-level simulation of expected shortfall: Confidence intervals, efficient simulation procedures, and high-performance computing. Ph.D. thesis, Northwestern University, Evanston, IL.

Lan, H., B. L. Nelson, J. Staum. 2007. Two-level simulations for risk management. S. Chick, C.-H. Chen, S. Henderson, E. Yücesan, eds. *Proc. 2007 INFORMS Simulation Soc. Res. Workshop*. INFORMS, Hanover, MD, 102–107.

Lee, S.-H. 1998. Monte Carlo computation of conditional expectation quantiles. Ph.D. thesis, Stanford University, Stanford, CA.

Searle, S. R., G. Casella, C. E. McCulloch. 1992. *Variance Components*. Wiley-Interscience, New York.

Staum, J. 2009. Monte Carlo computation in finance. P. L'Ecuyer, A. B. Owen, eds. *Monte Carlo and Quasi-Monte Carlo Methods 2008*. Springer-Verlag, New York, 19–42.

Steckley, S. G. 2006. Estimating the density of a conditional expectation. Ph.D. thesis, Cornell University, Ithaca, NY.

Steckley, S. G., S. G. Henderson. 2003. A kernel approach to estimating the density of a conditional expectation. S. Chick, P. J. Sánchez, D. Ferrin, D. J. Morrice, eds. *Proc. 2003 Winter Simulation Conf.* IEEE Press, Piscataway, NJ, 383–391.

Zouaoui, F., J. R. Wilson. 2003. Accounting for parameter uncertainty in simulation input modeling. *IIE Trans.* **35** 781–792.