$1\frac{1}{2}$-level Simulation for Estimating the Variance of a Conditional Expectation

Yunpeng Sun, Daniel W. Apley, and Jeremy Staum
Department of Industrial Engineering and Management Sciences
McCormick School of Engineering
Northwestern University
2145 Sheridan Road, Evanston, IL 60208-3119

August 18, 2009

Abstract

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 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 approaches a finite limit. 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 $1\frac{1}{2}$-level simulation. Because the optimal number of inner-level samples is often quite small, a $1\frac{1}{2}$-level simulation can avoid the heavy computational burden typically associated with two-level simulation.

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 $M := \mathrm{E}[X|Z]$ 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 may 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 may 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). Two-level simulation also has applications in decision theory (Brennan et al., 2007) and financial engineering (Staum, 2009).

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 $F_{X|Z=z}$ of X given $Z = z$ for any z, but that we can not sample directly from the distribution F_M of M.

If we were only interested in the mean μ, an ordinary one-level 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 $F_{X|Z=Z_k}$. 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 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 $F_M^{-1}(100p)$. Two-level 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 $F_{X|Z=Z_k}$.

In the simplest form of two-level 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 simulation provides an accurate estimator $\bar{X}_k := \sum_{j=1}^{n} X_{kj}/n$ of $M_k := \mathrm{E}[X|Z = Z_k]$. A straightforward estimator of the variance or 99th percentile of the distribution F_M is the variance or 99th percentile of the empirical distribution $\widehat{F_M}$ of $\bar{X}_1, \ldots, \bar{X}_K$, which is given by $\widehat{F_M}(y) = \sum_{k=1}^{K} 1\{\bar{X}_k \leq y\}/K$. For example, this estimator of the variance is

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

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. In particular, the estimator (1) is biased high, for the following reason. Define the conditional variance $V := \mathrm{Var}[X|Z]$. Then, for all k,

$$\mathrm{Var}\left[\bar{X}_k\right] = \mathrm{Var}\left[\mathrm{E}\left[\bar{X}_k|Z_k\right]\right] + \mathrm{E}\left[\mathrm{Var}\left[\bar{X}_k|Z_k\right]\right] = \sigma_M^2 + \mathrm{E}[V]/n > \sigma_M^2 \tag{2}$$

if V is nonzero. The presence of bias implies that both the outer- and inner-level sample sizes K and n of a two-level simulation must grow without bound to get the mean squared error to converge to zero. The literature on two-level 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, 2008; Lee, 1998).

Often, the finding is that the inner-level sample size may need to be quite large for a two-level 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 (2009). Gordy and Juneja (2008) reach a different conclusion in studying two-level 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 (2008) 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 call a nested simulation framework in which the inner-level sample size remains constant as the computational budget grows a $1\frac{1}{2}$-level simulation. 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 based on the assumption that the conditional variance V is the same for any scenario Z. In general, this assumption of constant conditional variance is not true—one of our contributions is to prove that the ANOVA approach provides an unbiased estimator even when the conditional variance depends on the scenario. Our second 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 approaches a finite limit 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 discuss how to choose a good inner-level sample size.

The rest of this paper is organized as follows. In Section 2 we present a general framework and our unbiased estimator. Section 3 is about the optimal inner-level sample size. A numerical example occupies Section 4. We give conclusions and research directions in Section 5. Some derivations are deferred to the appendix.

2 Derivation of Our 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

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

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: $\mathrm{E}[\tau\varepsilon] = \mathrm{E}[\tau \mathrm{E}[\varepsilon|\tau]] = 0 = \mathrm{E}[\tau]\mathrm{E}[\varepsilon]$.

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) looks like the standard one-way, random effects ANOVA model (Searle et al., 1992), so one might consider using standard ANOVA methods to estimate σ_M^2. However, the standard ANOVA model involves the assumption that the effects and errors are independent. We do not impose this assumption in model (3) because it is not generally true in simulation applications. In particular, the conditional error variance $V = \mathrm{Var}[\varepsilon|Z]$ is often strongly related to the conditional mean M and thus the effect. For example, in the queuing example mentioned in Section 1, scenarios that result in larger mean time in system may also have larger variability of the time in system.

There are other more general variance components models that allow the error variance $\mathrm{Var}[\varepsilon_{kj}]$ to depend on k (Searle et al., 1992, §4.6). However, the standard methods for estimating the variance components in these formulations require that one view $\sigma_k^2 = \mathrm{Var}[\varepsilon_{kj}]$ either as a constant that has no functional dependence on the effect τ_k or as a random variable that is stochastically independent of τ_k. Neither of these are reasonable assumptions for nested simulation applications in general. To handle the general situation in which the effects and errors are dependent, one might consider developing a variance components estimation method based on maximum likelihood estimation or minimum variance quadratic unbiased estimation (Searle et al., 1992), but this would require knowledge of a functional relationship that describes the mutual dependence between τ_k and ε_{kj}. For many applications, including those described in this paper, it would be unreasonable to assume that the analyst has such knowledge.

Fortunately, it is not necessary to have it: in the remainder of this section, we prove that the ordinary ANOVA estimators of the variance components are unbiased even when the effects and observation errors are dependent. 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

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

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

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

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

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

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

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

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

$$\widehat{\sigma_\varepsilon^2} = \frac{\mathrm{SS}_\varepsilon}{C-K} \quad \text{and} \quad \widehat{\sigma_M^2} = \frac{\mathrm{SS}_\tau - (K-1)\widehat{\sigma_\varepsilon^2}}{C - \sum_{i=1}^{K} n_i^2/C}. \qquad (8)$$

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

$$\widehat{\sigma_\varepsilon^2} = \frac{1}{K(n-1)}\sum_{k=1}^{K}\sum_{j=1}^{n}\left(X_{kj} - \bar{X}_k\right)^2 \quad \text{and} \quad \widehat{\sigma_M^2} = \frac{1}{K-1}\sum_{k=1}^{K}\left(\bar{X}_k - \bar{\bar{X}}\right)^2 - \frac{1}{n}\widehat{\sigma_\varepsilon^2}. \qquad (9)$$

Equations (8) and (9) are the usual ANOVA estimators for the standard model in which the effects and observation errors are assumed independent (Searle et al., 1992). The preceding derivation showed that these estimators are unbiased merely because the effects and the observation errors are uncorrelated. In contrast, the variance of $\widehat{\sigma_M^2}$ 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 assumes independence between the effects and observation errors, to facilitate testing of hypotheses related to variance components.

3 The Optimal Number of Inner Level Replicates

In this section, we study the variance of the estimator $\widehat{\sigma_M^2}$ 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 of 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 Appendix A shows that the variance of $\widehat{\sigma_M^2}$ is

$$\begin{aligned}\operatorname{Var}\left[\widehat{\sigma_M^2}\right] \quad = \quad & \frac{n}{C}\mathrm{E}\left[\tau^4\right] - \frac{n(C-3n)}{C(C-n)}\sigma_M^4 + \frac{2n}{C^2(C-n)}\sigma_\varepsilon^4 + \frac{2n(C+n)}{C^2(C-n)}\sigma_M^2\sigma_\varepsilon^2 \\ & + \frac{2}{C^2 n}\mathrm{E}\left[\varepsilon^4\right] + \frac{2(n^2+(C-4)n+3)}{C^2 n(n-1)}\mathrm{E}\left[V^2\right] + \frac{4C+2n}{C^2}\mathrm{E}\left[\tau^2\varepsilon^2\right] + \frac{4}{C^2}\mathrm{E}\left[\tau\varepsilon^3\right].\end{aligned} \tag{10}$$

Recall that $\varepsilon = X - M$ is the error associated with one observation of the effect $\tau = M - \mu$, so that ε and τ may be dependent, although uncorrelated: this is why $\mathrm{E}[\tau^2\varepsilon^2]$ and $\mathrm{E}[\tau\varepsilon^3]$ may be nonzero.

We see that minimizing Equation (10) over n requires knowing (or estimating) several cross-moments of the effects and errors. Results in Figure 2 suggest that in many problems, a nearly optimal choice of n can be found by using a much simpler approximation to Equation (10) that arises by supposing that the number K of scenarios is large. This case is especially worth considering because $K \to \infty$ is a necessary condition for $\operatorname{Var}\left[\widehat{\sigma_M^2}\right] \to 0$. We suppose that $K = C/n$ is so large that terms of order $o(1/K)$ are negligible. Then

$$\operatorname{Var}\left[\widehat{\sigma_M^2}\right] \approx \frac{1}{C}\left(n(\mathrm{E}\left[\tau^4\right] - \sigma_M^4) + \frac{2\mathrm{E}\left[V^2\right]}{n-1} + 4\mathrm{E}\left[\tau^2\varepsilon^2\right]\right). \tag{11}$$

Treating C as constant, differentiating Equation (11) with respect to n, and setting the result equal to zero, we find that the optimal inner-level sample size for a large budget C is approximately

$$n^* = 1 + \sqrt{\frac{2\mathrm{E}\left[V^2\right]}{\mathrm{E}\left[\tau^4\right] - \sigma_M^4}} = 1 + \sqrt{\frac{2\mathrm{E}\left[V^2\right]}{\sigma_M^4(\kappa-1)}} = 1 + \sqrt{\frac{2\left(\sigma_\varepsilon^4 + \operatorname{Var}\left[V\right]\right)}{\sigma_M^4(\kappa-1)}}, \tag{12}$$

where κ denotes the kurtosis of F_M or of the effect τ. By first differentiating Equation (10) with respect to n to find the optimal sample size n_C^* for a finite budget C, and then taking the limit as $C \to \infty$, one can show that n^* in Equation (12) is the asymptotically optimal inner-level sample size. The remarkable finding that it is a finite constant and does not grow with the budget C is the basis for the phrase "$1\frac{1}{2}$-level simulation."

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 $\operatorname{Var}[V]$ of inner-level variance, the outer-level variance σ_M^2, and the outer-level kurtosis κ. 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 and the ratio of $\mathrm{E}[V^2] = \mathrm{E}[\operatorname{Var}[X|Z]^2]$ to $\sigma_M^4 = \operatorname{Var}[\mathrm{E}[X|Z]]^2$. Figure 1 shows how n^* depends on the outer-level kurtosis κ and on the fourth root of this ratio; if the inner-level variance $V = \operatorname{Var}[X|Z]$ does not depend on the scenario, then $\operatorname{Var}[V] = 0$ 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 1 are 3, 9 and 73.8, respectively—a wide range

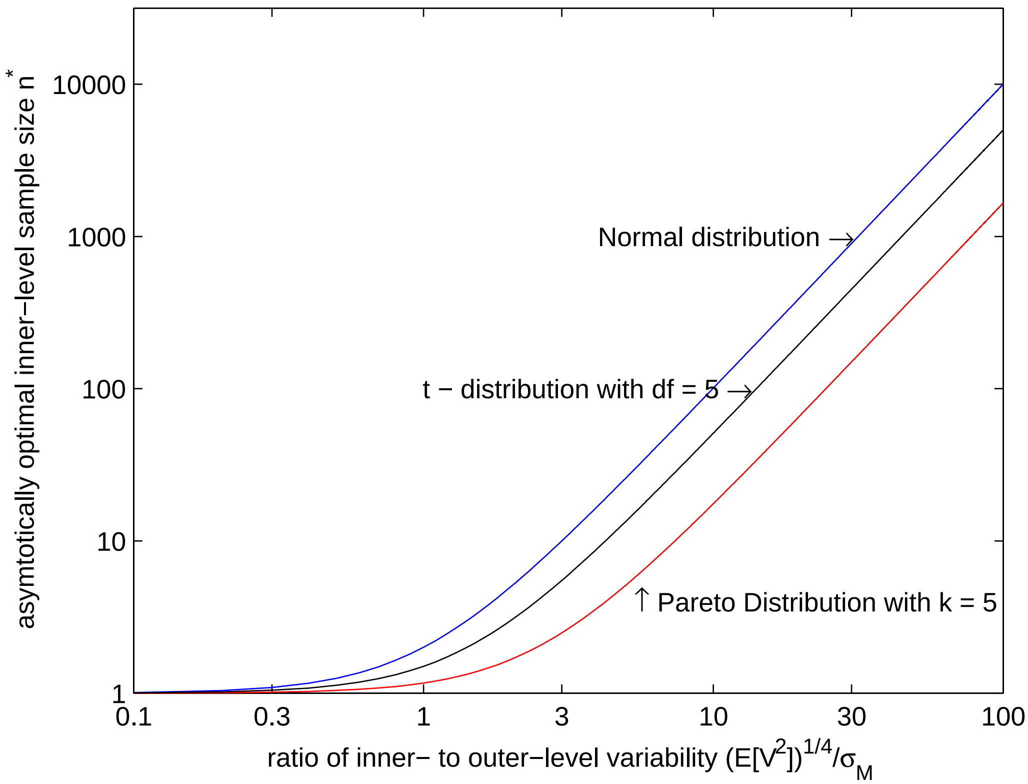

Figure 1: The inner-level sample size n^* that is asymptotically optimal for large computational budgets, given by Equation (12), versus a ratio of inner- to outer-level variability, for three different outer-level distributions.

of values. For many simulation problems encountered in practice, the outer-level kurtosis will be between 3 and 73.8, and hence n^* will lie somewhere between the curves shown for the normal and Pareto distributions in Figure 1. For $\kappa \leq 73.8$ and $\mathrm{E}[V^2] \leq \sigma_M^4$, n^* must be rounded up to 2, which is the smallest inner-level sample size that supports unbiased estimation of σ_M^2. For $\kappa \leq 73.8$ and $\mathrm{E}[V^2]^{1/4}/\sigma_M \leq 3$, $n^* \leq 10$; for $\mathrm{E}[V^2]^{1/4}/\sigma_M \leq 10$, $n^* \leq 100$. 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 may wonder how large C must be before n^* is nearly optimal for a finite budget C. Figure 2 answers this question for a special case that fits a standard ANOVA framework: errors and effects are independent, and effects are normally distributed. We obtained the optimal inner-level sample size for a finite 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

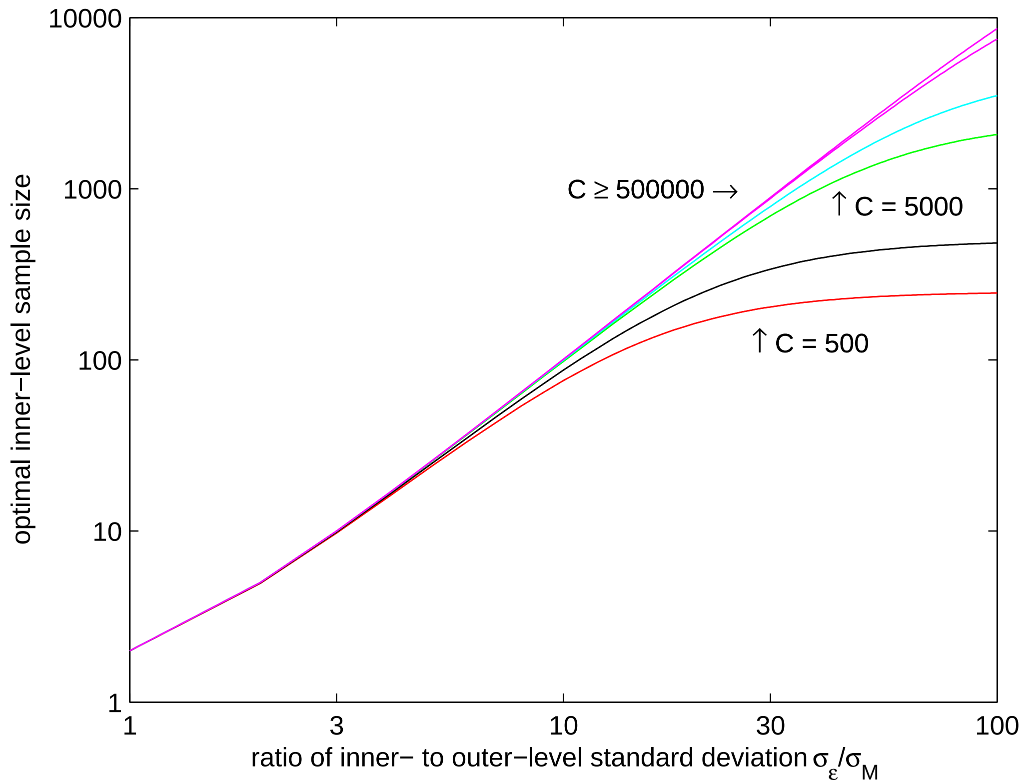

Figure 2: The inner-level sample size that minimizes the estimator variance in Equation (10) for a fixed computational budget C when errors and effects are independent and effects are normally distributed, versus the ratio of the standard deviations of errors and effects.

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 guesses for or estimates of unknown quantities in Equation (12). Appendix B presents a method for estimating n^* based on simulation output. In our experience, it can be difficult to estimate n^* accurately with this method using a pilot simulation that is computationally inexpensive compared to the main simulation whose budget is C. In the example treated in the next section, we get a good estimate of n^* from the output of a simulation with a large budget and large inner-level sample size. Thus, 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 may be able to estimate n^* well from the output of those experiments. By improving upon the method in Appendix B, 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.

4 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, similar to one used by Baysal et al. (2008), 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. Appendix C provides more details about the example. Here we only give a high-level description and explanation of how it fits in our ANOVA framework. 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 of one European put option and one European call option 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. 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 variance depends on details of the delta-hedging strategy, such as how many times s the portfolio is re-balanced before time T: at times $t_0 = 0, t_1, \ldots, t_{s-1}$, the portfolio is adjusted by changing the number of shares of stock to equal the current sensitivity of the original portfolio to the stock price.

The P&L is a function of the path of stock prices at times $t_0 = 0, t_1, \ldots, t_s = T$. This path is the scenario Z in our ANOVA framework, and the P&L that results from this path is the conditional expectation $M = \mathrm{E}[X|Z]$, where the random variable X is given in Appendix C. 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. Figure 3 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, as explained in Appendix C. Each panel of Figure 3 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.

Figure 4 illustrates the benefit of $1\frac{1}{2}$-level simulation by showing how the variance of the ANOVA estimator $\widehat{\sigma_M^2}$ 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 $\widehat{\sigma_M^2}$. 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 Appendix B, we estimated the asymptotically optimal inner-level sample size n^* of Equation (12) by $\widehat{n^*} = 45$. We then ran nested simulations with different inner-level sample sizes n to see how the variance of $\widehat{\sigma_M^2}$ with $n = \widehat{n^*}$ compares to the variance with other choices of n. This exercise also demonstrated good agreement of the formula for $\mathrm{Var}[\widehat{\sigma_M^2}]$ in Equation (10), where estimates were substituted for unknown quantities, with the direct estimates of $\mathrm{Var}[\widehat{\sigma_M^2}]$ based on macro-replications. The numerical results indicate that $\widehat{n^*} = 45$ is indeed nearly optimal for the computational budgets considered here. These results provide some validation for our analysis of $\mathrm{Var}[\widehat{\sigma_M^2}]$ 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 1600. Figure 4 shows that using $n = 1600$ instead of $n = 45$ makes the variance of $\widehat{\sigma_M^2}$ 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_M^2 by $\widehat{\sigma_M^2}$ that is attained with budget $C = 800,000$ and $n = 45$, if we were to use $n = 1600$ then we would require a budget of over 10 million.

5 Conclusions and Research Directions

Our principal findings are twofold. First, the ANOVA estimator $\widehat{\sigma_M^2}$ 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 approach a finite limit 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). 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 purpose of estimating the variance of the conditional expectation. The smaller sample size can greatly increase computational efficiency.

We believe that 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 $\sigma_M^2 = \mathrm{Var}[M]$ and hence $\mathrm{E}[M^2]$ if the inner-level sample size $n \geq 2$. Likewise, it is not hard to show how to construct an

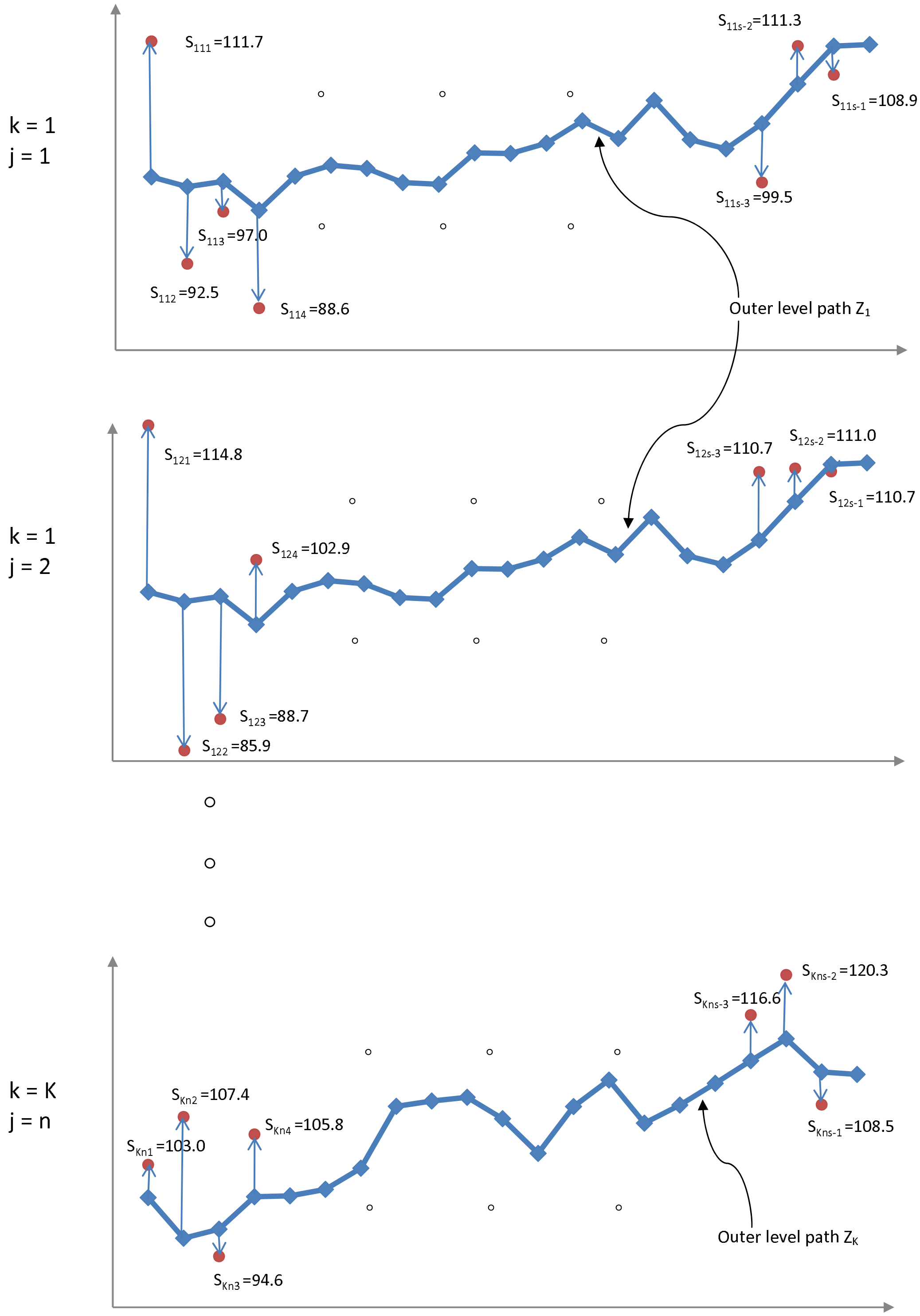

Figure 3: Illustration of nested simulation in the delta-hedging example.

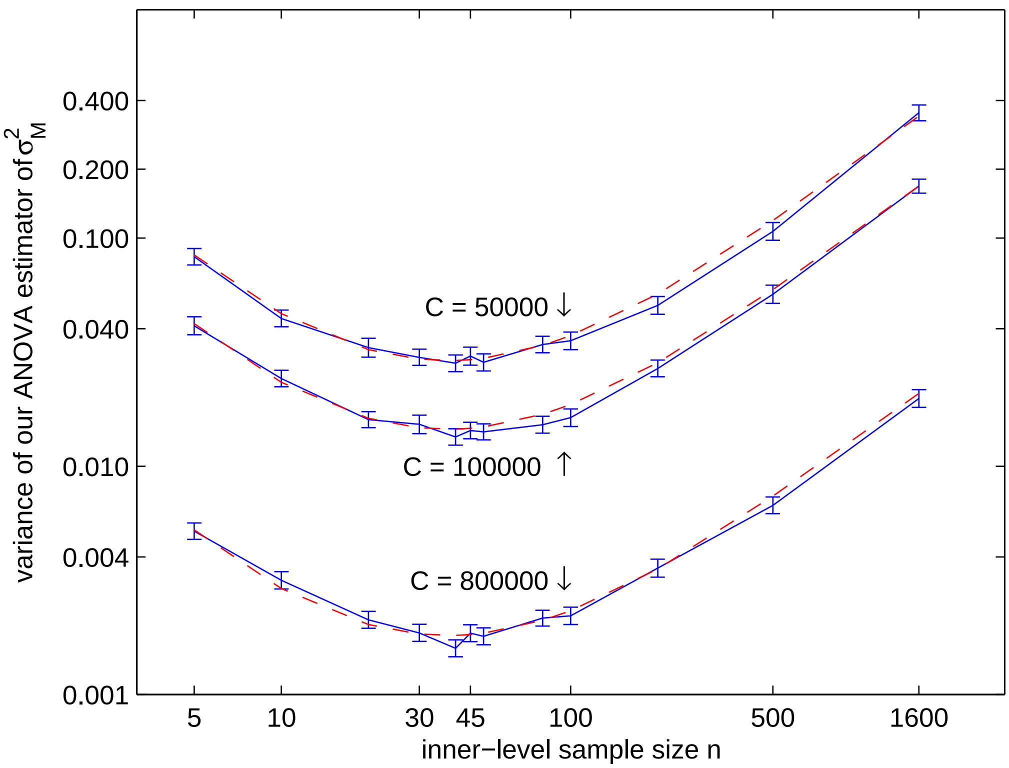

Figure 4: Variance of the ANOVA estimator $\widehat{\sigma_M^2}$ in the delta-hedging example, as a function of the inner-level sample size given a fixed computational budget, for three different computational budgets C. The solid curves give point estimates of the variance and the error bars are 95% confidence intervals for the variance. The dashed curves represent the formula for the variance given in Equation (10), with estimates substituted for unknown quantities.

unbiased estimator of $\mathrm{E}[M^m]$ if $n \geq m$. 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 $\mathrm{E}[f(M)]$. 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 1 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 $\widehat{F_M}$.

Acknowledgments

This research was partially supported by National Science Foundation grants number DMI-0555485 and CMMI-0758557. We thank Paul Glasserman and Shane Henderson for suggesting the research direction involving the Taylor expansion.

References

Baysal, R. E., Nelson, B. L., Staum, J., 2008. Response surface methodology for hedging and trading strategies. In: Mason, S. J., Hill, R. R., Mönch, L., Rose, O., Jefferson, T., Fowler, J. W. (Eds.), Proceedings of the 2008 Winter Simulation Conference. IEEE Press, Piscataway, N. J., pp. 629–637.

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

Chick, S. E., 2001. Input distribution selection for simulation experiments: accounting for input uncertainty. Operations Research 49 (5), 744–758.

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

Gordy, M. B., Juneja, S., April 2008. Nested simulation in portfolio risk measurement. Finance and Economics Discussion Series 2008-21, Federal Reserve Board.

Lan, H., 2009. Two-level simulation of expected shortfall: Confidence intervals, efficient simulation procedures, and high-performance computing. Ph.D. thesis, Northwestern University.

Lan, H., Nelson, B. L., Staum, J., 2007. Two-level simulations for risk management. In: Chick, S., Chen, C.-H., Henderson, S., Yücesan, E. (Eds.), Proceedings of the 2007 INFORMS Simulation Society Research Workshop. pp. 102–107.

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

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

Staum, J., 2009. Monte Carlo computation in finance. In: L'Ecuyer, P., Owen, A. B. (Eds.), Monte Carlo and Quasi-Monte Carlo 2008. Springer-Verlag, New York, forthcoming.

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

A Variance of the ANOVA estimator

The variance of the estimator defined in Equation (9) is

$$\mathrm{Var}\left[\widehat{\sigma^2_M}\right] = \frac{\mathrm{Var}[\mathrm{SS}_\tau]}{n^2(K-1)^2} + \frac{\mathrm{Var}[\mathrm{SS}_\varepsilon]}{n^2K^2(n-1)^2} - \frac{2\mathrm{Cov}[\mathrm{SS}_\tau, \mathrm{SS}_\varepsilon]}{n^2K(K-1)(n-1)}.$$

Expressions for the three terms $\mathrm{Var}[\mathrm{SS}_\varepsilon]$, $\mathrm{Cov}[\mathrm{SS}_\tau, \mathrm{SS}_\varepsilon]$, and $\mathrm{Var}[\mathrm{SS}_\tau]$ appear in Equations (13), (15) and (23) below. Summing them yields Equation (10).

Using Equation (7) in the third line of the following derivation,

$$\begin{aligned}
\mathrm{Var}[\mathrm{SS}_\varepsilon] &= \sum_{k=1}^{K} \mathrm{Var}\left[\sum_{j=1}^{n}(\varepsilon_{kj} - \bar{\varepsilon}_k)^2\right] \\
&= \sum_{k=1}^{K}\left\{\mathrm{E}\left[\left(\sum_{j=1}^{n}(\varepsilon_{kj} - \bar{\varepsilon}_k)^2\right)^2\right] - \left(\mathrm{E}\left[\sum_{j=1}^{n}(\varepsilon_{kj} - \bar{\varepsilon}_k)^2\right]\right)^2\right\} \\
&= \sum_{k=1}^{K}\mathrm{E}\left[\left(\sum_{j=1}^{n}\frac{n-1}{n}\varepsilon_{kj}^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{i'\neq i}\varepsilon_{ki}\varepsilon_{ki'}\right)^2\right] - \sum_{k=1}^{K}(n-1)^2\sigma_\varepsilon^4 \\
&= K\left(\frac{(n-1)^2}{n}\mathrm{E}[\varepsilon^4] + \left(\frac{2(n-1)}{n} + \frac{(n-1)^3}{n}\right)\mathrm{E}[V^2] + 0\right) - K(n-1)^2\sigma_\varepsilon^4 \\
&= \frac{K(n-1)^2}{n}\mathrm{E}[\varepsilon^4] + \frac{K(n-1)(n^2-2n+3)}{n}\mathrm{E}[V^2] - K(n-1)^2\sigma_\varepsilon^4. \qquad (13)
\end{aligned}$$

Next, the covariance $\mathrm{Cov}[\mathrm{SS}_\tau, \mathrm{SS}_\varepsilon] = \mathrm{E}[\mathrm{SS}_\tau \mathrm{SS}_\varepsilon] - \mathrm{E}[\mathrm{SS}_\tau]\mathrm{E}[\mathrm{SS}_\varepsilon]$. Equations (6) and (7) say that $\mathrm{E}[\mathrm{SS}_\tau] = (K-1)(n\sigma_M^2 + \sigma_\varepsilon^2)$ and $\mathrm{E}[\mathrm{SS}_\varepsilon] = K(n-1)\sigma_\varepsilon^2$. The remaining term

$$\begin{aligned}
\mathrm{E}[\mathrm{SS}_\tau \mathrm{SS}_\varepsilon] &= \mathrm{E}\left[\sum_{k=1}^{K} n((\tau_k - \bar{\tau}) + (\bar{\varepsilon}_k - \bar{\bar{\varepsilon}}))^2 \sum_{k'=1}^{K}\sum_{j=1}^{n} (\varepsilon_{k'j} - \bar{\varepsilon}_{k'})^2\right] \\
&= \sum_{k=1}^{K}\sum_{j=1}^{n} n\mathrm{E}\left[\left((\tau_k - \bar{\tau}) + (\bar{\varepsilon}_k - \bar{\bar{\varepsilon}})\right)^2 (\varepsilon_{kj} - \bar{\varepsilon}_k)^2\right] \\
&\quad + \sum_{k=1}^{K}\sum_{k'\neq k}\sum_{j=1}^{n} n\mathrm{E}\left[(\tau_k - \bar{\tau} + \bar{\varepsilon}_k - \bar{\bar{\varepsilon}})^2\right]\mathrm{E}\left[(\varepsilon_{k'j} - \bar{\varepsilon}_{k'})^2\right] \\
&= Kn^2\mathrm{E}\left[\left((\tau_1 - \bar{\tau}) + (\bar{\varepsilon}_1 - \bar{\bar{\varepsilon}})\right)^2 (\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] + \frac{K-1}{K}\mathrm{E}[\mathrm{SS}_\tau]\mathrm{E}[\mathrm{SS}_\varepsilon].
\end{aligned}$$

Therefore $\mathrm{Cov}[\mathrm{SS}_\tau, \mathrm{SS}_\varepsilon]$ is

$$Kn^2\mathrm{E}\left[\left((\tau_1 - \bar{\tau}) + (\bar{\varepsilon}_1 - \bar{\bar{\varepsilon}})\right)^2 (\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] - (K-1)n(n-1)\sigma_\varepsilon^2\sigma_M^2 - (K-1)(n-1)\sigma_\varepsilon^4. \quad (14)$$

It remains to calculate $\mathrm{E}\left[\left((\tau_1 - \bar{\tau}) + (\bar{\varepsilon}_1 - \bar{\bar{\varepsilon}})\right)^2 (\varepsilon_{11} - \bar{\varepsilon}_1)^2\right]$, which equals

$$\mathrm{E}\left[(\tau_1 - \bar{\tau})^2(\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] + 2\mathrm{E}\left[(\tau_1 - \bar{\tau})(\bar{\varepsilon}_1 - \bar{\bar{\varepsilon}})(\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] + \mathrm{E}\left[(\bar{\varepsilon}_1 - \bar{\bar{\varepsilon}})^2(\varepsilon_{11} - \bar{\varepsilon}_1)^2\right].$$

Next we calculate each of these three terms, making use of the facts that the effects and errors have zero mean and are uncorrelated. We also use the idea that showed $\mathrm{E}[(\varepsilon_{11} - \bar{\varepsilon}_1)^2] = \sigma_\varepsilon^2(n-1)/n$ in the derivation of Equation (7). First,

$$\begin{aligned}
\mathrm{E}\left[(\tau_1 - \bar{\tau})^2(\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] &= \left(1 - \frac{1}{K}\right)^2 \mathrm{E}\left[\tau_1^2(\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] + \frac{1}{K^2}\sum_{k=2}^{K}\mathrm{E}\left[\tau_k^2\right]\mathrm{E}\left[(\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] \\
&= \frac{(K-1)^2(n-1)}{K^2 n}\mathrm{E}\left[\tau^2\varepsilon^2\right] + \frac{(K-1)(n-1)}{K^2 n}\sigma_M^2\sigma_\varepsilon^2.
\end{aligned}$$

Second,

$$\mathrm{E}\left[(\tau_1 - \bar{\tau})(\bar{\varepsilon}_1 - \bar{\bar{\varepsilon}})(\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] = \frac{(K-1)^2(n-1)}{K^2 n}\mathrm{E}\left[\tau\varepsilon^3\right].$$

Third, using $\bar{\varepsilon}_1 - \bar{\bar{\varepsilon}} = \varepsilon_1(K-1)/K - \sum_{k=2}^{K}\varepsilon_k/K$,

$$\begin{aligned}
\mathrm{E}\left[(\bar{\varepsilon}_1 - \bar{\bar{\varepsilon}})^2(\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] &= \left(1 - \frac{1}{K}\right)^2 \mathrm{E}\left[\bar{\varepsilon}_1^2(\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] + \frac{1}{K^2}\sum_{k=2}^{K}\mathrm{E}\left[\bar{\varepsilon}_k^2\right]\mathrm{E}\left[(\varepsilon_{11} - \bar{\varepsilon}_1)^2\right] \\
&= \frac{(K-1)^2}{K^2}\left(\mathrm{E}[\bar{\varepsilon}_1^2\varepsilon_{11}^2] - 2\mathrm{E}[\varepsilon_{11}(\bar{\varepsilon}_1)^3] + \mathrm{E}[\bar{\varepsilon}_1^4]\right) + \frac{(K-1)(n-1)}{K^2 n^2}\sigma_\varepsilon^4.
\end{aligned}$$

Because $\mathrm{E}[\bar{\varepsilon}_1^2\varepsilon_{11}^2] - 2\mathrm{E}[\varepsilon_{11}(\bar{\varepsilon}_1)^3] + \mathrm{E}[\bar{\varepsilon}_1^4]$ equals

$$\frac{\mathrm{E}[\varepsilon^4]}{n} + \frac{(n-1)\mathrm{E}[V^2]}{n^2} - \frac{2\mathrm{E}[\varepsilon^4]}{n^3} - \frac{6(n-1)\mathrm{E}[V^2]}{n^3} + \frac{\mathrm{E}[\varepsilon^4]}{n^3} + \frac{3(n-1)\mathrm{E}[V^2]}{n^3} + \frac{(K-1)(n-1)\sigma_\varepsilon^4}{K^2 n^2},$$

we find that $\mathrm{E}\left[(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2(\varepsilon_{11}-\bar{\varepsilon}_1)^2\right]$ is

$$\frac{(K-1)^2(n-1)}{K^2n^3}\mathrm{E}[\varepsilon^4]+\frac{(K-1)^2(n-1)(n-3)}{K^2n^3}\mathrm{E}[V^2]+\frac{(K-1)(n-1)}{K^2n^2}\sigma_\varepsilon^4.$$

Substituting these three terms into (14),

$$\begin{aligned}\mathrm{Cov}\left[\mathrm{SS}_\tau,\mathrm{SS}_\varepsilon\right] &= \frac{n(n-1)(K-1)^2}{K}\mathrm{E}\left[\tau^2\varepsilon^2\right]-\frac{n(n-1)(K-1)^2}{K}\sigma_\varepsilon^2\sigma_M^2\\ &\quad+\frac{2n(n-1)(K-1)^2}{K}\mathrm{E}\left[\tau\varepsilon^3\right]+\frac{(n-1)(K-1)^2}{nK}\mathrm{E}\left[\varepsilon^4\right]\\ &\quad+\frac{(n-3)(n-1)(K-1)^2}{Kn}\mathrm{E}\left[V^2\right]-\frac{(n-1)(K-1)^2}{K}\sigma_\varepsilon^4. \qquad (15)\end{aligned}$$

Finally, we come to $\mathrm{Var}[\mathrm{SS}_\tau]$. From

$$\mathrm{SS}_\tau=\sum_{k=1}^K n\left(\bar{X}_k-\bar{\bar{X}}\right)^2=\sum_{k=1}^K n\left(\tau_k-\bar{\tau}+\bar{\varepsilon}_k-\bar{\bar{\varepsilon}}\right)^2,$$

it follows that

$$\begin{aligned}\mathrm{Var}[\mathrm{SS}_\tau] &= n^2\sum_{k=1}^K \mathrm{Var}\left[(\tau_k-\bar{\tau}+\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})^2\right]\\ &\quad+n^2\sum_{k=1}^K\sum_{k'\neq k}\mathrm{Cov}\left[((\tau_k-\bar{\tau})+(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}}))^2,((\tau_{k'}-\bar{\tau})+(\bar{\varepsilon}_{k'}-\bar{\bar{\varepsilon}}))^2\right] \qquad (16)\end{aligned}$$

The first term is

$$Kn^2\left(\mathrm{E}[(\tau_1-\bar{\tau}+\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^4]-\mathrm{E}[(\tau_1-\bar{\tau}+\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2]^2\right). \qquad (17)$$

In (17), because effects and errors are uncorrelated, $\mathrm{E}[(\tau_1-\bar{\tau}+\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2]$ is

$$\mathrm{E}\left[\left(\left(\tau_1-\frac{1}{K}\sum_{i=1}^K\tau_i\right)+\left(\bar{\varepsilon}_1-\frac{1}{K}\sum_{i=1}^K\bar{\varepsilon}_i\right)\right)^2\right]=\left(\frac{K-1}{K}\right)\sigma_M^2+\frac{K-1}{Kn}\sigma_\varepsilon^2. \qquad (18)$$

The other expectation in (17) is

$$\begin{aligned}\mathrm{E}[(\tau_1-\bar{\tau}+\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^4] &= \mathrm{E}[\left((\tau_1-\bar{\tau})^2+2(\tau_1-\bar{\tau})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})+(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})\right)^2]\\ &= \mathrm{E}[(\tau_1-\bar{\tau})^4]+4\mathrm{E}[(\tau_1-\bar{\tau})^3(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})]+6\mathrm{E}[(\tau_1-\bar{\tau})^2(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2]\\ &\quad+4\mathrm{E}[(\tau_1-\bar{\tau})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^3]+\mathrm{E}[(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^4]. \qquad (19)\end{aligned}$$

We now calculate each of these expectations. The easiest one is

$$\mathrm{E}[(\tau_1-\bar{\tau})^3(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})]=\mathrm{E}[\mathrm{E}[(\tau_1-\bar{\tau})^3(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})|Z_1]]=\mathrm{E}[\mathrm{E}[(\tau_1-\bar{\tau})^3|Z_1]\mathrm{E}[(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})|Z_1]]=0.$$

Next, $\mathrm{E}[(\tau_1-\bar{\tau})^4]$ is

$$\begin{aligned}
&\mathrm{E}\left[\left(\left(1-\frac{1}{K}\right)\tau_1-\frac{1}{K}\sum_{k=2}^{K}\tau_k\right)^4\right]\\
&= \mathrm{E}\Bigg[\left(1-\frac{1}{K}\right)^4\tau_1^4+\frac{1}{K^4}\sum_{k=2}^{K}\tau_k^4+\left(1-\frac{1}{K}\right)^2\frac{4}{K^2}\sum_{k=2}^{K}\tau_1^2\tau_k^2\\
&\quad+\frac{4}{K^4}\sum_{k=2}^{K}\sum_{s>k}^{K}\tau_k^2\tau_s^2+\left(1-\frac{1}{K}\right)^2\frac{2}{K^2}\sum_{k=2}^{K}\tau_1^2\tau_k^2+\frac{2}{K^4}\sum_{k=2}^{K}\sum_{s>k}^{K}\tau_k^2\tau_s^2\Bigg]\\
&= \left(\left(1-\frac{1}{K}\right)^4+\frac{K-1}{K^4}\right)\mathrm{E}[\tau^4]+\frac{6(K-1)^3}{K^4}\mathrm{E}[\tau^2]^2+\frac{3(K-1)(K-2)}{K^4}\mathrm{E}[\tau^2]^2\\
&= \frac{(K-1)^4+(K-1)}{K^4}\mathrm{E}[\tau^4]+\frac{3(K-1)(2K-3)}{K^3}\sigma_M^4.
\end{aligned}$$

Using similar reasoning,

$$\begin{aligned}
\mathrm{E}[(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^4] &= \frac{(K-1)^4+(K-1)}{K^4}\mathrm{E}[\bar{\varepsilon}_1^4]+\frac{3(K-1)(2K-3)}{K^3}\sigma_{\bar{\varepsilon}}^4\\
&= \frac{(K-1)^4+(K-1)}{K^4n^3}\mathrm{E}[\varepsilon^4]+\frac{3(n-1)((K-1)^4+(K-1))}{K^4n^3}E[V^2]\\
&\quad+\frac{3(K-1)(2K-3)}{K^3n^2}\sigma_\varepsilon^4.
\end{aligned}$$

Next, $\mathrm{E}\left[(\tau_1-\bar{\tau})^2(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2\right]$ is

$$\begin{aligned}
&\mathrm{E}\left[\left(\left(1-\frac{1}{K}\right)^2\tau_1^2+\frac{1}{K^2}\sum_{k=2}^{K}\tau_k^2\right)\left(\left(1-\frac{1}{K}\right)^2\bar{\varepsilon}_1^2+\frac{1}{K^2}\sum_{k=2}^{K}\bar{\varepsilon}_k^2\right)\right]\\
&= \left(1-\frac{1}{K}\right)^4\mathrm{E}\left[\tau_1^2\bar{\varepsilon}_1^2\right]+\frac{1}{K^2}\left(1-\frac{1}{K}\right)^2\sum_{k=2}^{K}\mathrm{E}\left[\tau_k^2\right]\mathrm{E}\left[\bar{\varepsilon}_1^2\right]\\
&\quad+\frac{1}{K^2}\left(1-\frac{1}{K}\right)^2\sum_{k=2}^{K}\mathrm{E}\left[\tau_1^2\right]\mathrm{E}\left[\bar{\varepsilon}_k^2\right]+\frac{1}{K^4}\sum_{k=2}^{K}\sum_{s\neq k}^{K}\mathrm{E}\left[\tau_k^2\bar{\varepsilon}_s^2\right]+\frac{1}{K^4}\sum_{k=2}^{K}\mathrm{E}\left[\tau_k^2\bar{\varepsilon}_k^2\right]\\
&= \left(\left(\frac{K-1}{K}\right)^4+\frac{K-1}{K^4}\right)\frac{1}{n}\mathrm{E}\left[\tau^2\varepsilon^2\right]+\frac{1}{n}\left(\frac{2(K-1)^3}{K^4}+\frac{(K-1)(K-2)}{K^4}\right)\mathrm{E}\left[\tau^2\right]\mathrm{E}\left[\varepsilon^2\right]\\
&= \frac{(K-1)^4+(K-1)}{K^4n}\mathrm{E}\left[\tau^2\varepsilon^2\right]+\frac{2(K-1)^3+(K-1)(K-2)}{K^4n}\sigma_M^2\sigma_\varepsilon^2.
\end{aligned}$$

The last remaining expectation in Equation (19) is $\mathrm{E}\left[(\bar{\tau}_1-\bar{\bar{\tau}})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^3\right]$, which equals

$$\mathrm{E}\left[\left(1-\frac{1}{K}\right)\tau_1\left(\bar{\varepsilon}_1\left(1-\frac{1}{K}\right)-\frac{1}{K}\sum_{k=2}^{K}\bar{\varepsilon}_k\right)^3\right]-\mathrm{E}\left[\sum_{k=2}^{K}\frac{\tau_k}{K}(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^3\right].$$

The latter term is

$$\mathrm{E}\left[\sum_{k=2}^{K}\frac{\tau_k}{K}\left(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}}\right)^3\right]=-\frac{1}{K^4}\mathrm{E}\left[\sum_{k=2}\tau_1\bar{\varepsilon}_1^3\right]=-\frac{K-1}{K^4n^2}\mathrm{E}\left[\tau\varepsilon^3\right]$$

and the first term is

$$\begin{aligned}
&\mathrm{E}\left[\left(\left(1-\frac{1}{K}\right)\tau_1\left(\bar{\varepsilon}_1\left(1-\frac{1}{K}\right)-\frac{1}{K}\sum_{k=2}^{K}\bar{\varepsilon}_k\right)^3\right]\\
&= \mathrm{E}\Bigg[\left(1-\frac{1}{K}\right)\tau_1\Bigg(\bar{\varepsilon}_1^2\left(1-\frac{1}{K}\right)^2-\frac{2(K-1)}{K^2}\sum_{k=2}^{K}\bar{\varepsilon}_1\bar{\varepsilon}_k\\
&\quad+\frac{1}{K^2}\sum_{k=2}^{K}\bar{\varepsilon}_k^2+\frac{1}{K^2}\sum_{k=2}^{K}\sum_{k'\neq k}^{K}\bar{\varepsilon}_k\bar{\varepsilon}_{k'}\Bigg)\left(\bar{\varepsilon}_1\left(1-\frac{1}{K}\right)-\frac{1}{K}\sum_{k=2}^{K}\bar{\varepsilon}_k\right)\Bigg]\\
&= \mathrm{E}\left[\left(1-\frac{1}{K}\right)\tau_1\left(\left(1-\frac{1}{K}\right)^3\bar{\varepsilon}_1^3+\frac{4}{K^2}\left(1-\frac{1}{K}\right)\sum_{k=2}^{K}\bar{\varepsilon}_k^2\bar{\varepsilon}_1-\frac{1}{K^3}\sum_{k=2}^{K}\varepsilon_k^3\right)\right]\\
&= \left(1-\frac{1}{K}\right)^4\mathrm{E}[\tau\bar{\varepsilon}^3]=\frac{(K-1)^4}{K^4n^2}\mathrm{E}[\tau\varepsilon^3].
\end{aligned}$$

Summing these two terms,

$$\mathrm{E}\left[(\bar{\tau}_1-\bar{\bar{\tau}})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^3\right]=\frac{(K-1)^4+(K-1)}{K^4n^2}\mathrm{E}[\tau\varepsilon^3].$$

This completes the evaluation of Equation (19). Having evaluated Equations (18) and (19), we substitute them into (17). We thus find that $n^2\sum_{k=1}^{K}\mathrm{Var}\left[(\tau_k-\bar{\tau}+\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})^2\right]$, the first term of Equation (16), is

$$\begin{aligned}
&\frac{n^2((K-1)^4+(K-1))}{K^3}\mathrm{E}[\tau^4]-\frac{n^2(K-1)(K^2-7K+9)}{K^2}\sigma_M^4\\
&\quad-\frac{(K-1)(K^2-7K+9)}{K^2}\sigma_\varepsilon^4-\frac{2n(K-1)(K^2-7K+9)}{K^2}\sigma_M^2\sigma_\varepsilon^2\\
&\quad+\frac{(K-1)^4+(K-1)}{K^3n}\mathrm{E}[\varepsilon^4]+\frac{3(n-1)((K-1)^4+(K-1))}{K^3n}E[V^2]\\
&\quad+\frac{4(K-1)^4+4(K-1)}{K^3}\mathrm{E}[\tau\varepsilon^3]+\frac{6n((K-1)^4+(K-1))}{K^3}\mathrm{E}[\tau^2\varepsilon^2]. \qquad (20)
\end{aligned}$$

Next we compute $Kn^2\sum_{k=2}^{K}\mathrm{Cov}\left[((\tau_1-\bar{\tau})+(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}}))^2,((\tau_k-\bar{\tau})+(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}}))^2\right]$, the second term in Equation (16). For any k, $\mathrm{E}\left[(\tau_k-\bar{\tau}+\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})^2\right]$ equals Equation (18). Therefore, to calculate the covariance, it remains only to calculate

$$\begin{aligned}
&\mathrm{E}\left[((\tau_1-\bar{\tau})+(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}}))^2((\tau_k-\bar{\tau})+(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}}))^2\right]= \qquad (21)\\
&\quad\mathrm{E}\left[\left((\tau_1-\bar{\tau})^2+2(\tau_1-\bar{\tau})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})+(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2\right)\left((\tau_k-\bar{\tau})^2+2(\tau_k-\bar{\tau})(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})+(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})^2\right)\right]
\end{aligned}$$

for $k>1$. Because $\mathrm{E}[(\tau_1-\bar{\tau})^2(\tau_k-\bar{\tau})(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})]$ equals

$$\mathrm{E}[\mathrm{E}[(\tau_1-\bar{\tau})^2(\tau_k-\bar{\tau})(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})|Z_k]] = \mathrm{E}[(\tau_1-\bar{\tau})^2(\tau_k-\bar{\tau})\mathrm{E}[(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})|Z_k]] = 0,$$

Equation (21) equals

$$\begin{aligned}
&\mathrm{E}[(\tau_1-\bar{\tau})^2(\tau_k-\bar{\tau})^2] + \mathrm{E}[(\tau_k-\bar{\tau})^2(\varepsilon_1-\bar{\bar{\varepsilon}})^2] \\
&\quad + 4\mathrm{E}[(\tau_1-\bar{\tau})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})(\tau_k-\bar{\tau})(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})] + 2\mathrm{E}[(\tau_k-\bar{\tau})(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2] \\
&\quad + 2\mathrm{E}[(\tau_1-\bar{\tau})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})^2] + \mathrm{E}[(\tau_1-\bar{\tau})^2(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})^2] + \mathrm{E}[(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})^2]. \quad (22)
\end{aligned}$$

We next calculate each term in Equation (22). First,

$$\begin{aligned}
&\mathrm{E}\left[(\tau_1-\bar{\tau})^2(\tau_k-\bar{\tau})^2\right] \\
&= \mathrm{E}\Big[\tau_1^2\tau_k^2 - 2\tau_1\tau_k^2\bar{\tau} - 2\tau_1^2\tau_k\bar{\tau} + \tau_k^2(\bar{\tau})^2 + \tau_1^2(\bar{\tau})^2 + 4\tau_1\tau_k(\bar{\tau})^2 - 2\tau_k(\bar{\tau})^3 - 2\tau_1(\bar{\tau})^3 + (\bar{\tau})^4\Big] \\
&= \sigma_M^4 - \frac{4}{K}\sigma_M^4 + \frac{2}{K^2}\mathrm{E}[\tau^4] + \frac{2(K-1)}{K^2}\sigma_M^4 + \frac{8}{K^2}\sigma_M^4 \\
&\quad - \frac{4}{K^3}\mathrm{E}[\tau^4] - \frac{12(K-1)}{K^3}\sigma_M^4 + \frac{1}{K^3}\mathrm{E}[\tau^4] + \frac{3(K-1)}{K^3}\sigma_M^4 \\
&= \frac{2K-3}{K^3}\mathrm{E}[\tau^4] + \frac{K^3-2K^2-3K+9}{K^3}\sigma_M^4.
\end{aligned}$$

In a similar way, $\mathrm{E}[(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})^2]$ is

$$\frac{2K-3}{K^3n^3}\mathrm{E}[\varepsilon^4] + \frac{3(2K-3)(n-1)}{K^3n^3}\mathrm{E}[V^2] + \frac{K^3-2K^2-3K+9}{K^3n^2}\sigma_\varepsilon^4.$$

Next, $\mathrm{E}\left[(\tau_1-\bar{\tau})^2(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})^2\right]$ and $\mathrm{E}\left[(\tau_k-\bar{\tau})^2(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2\right]$ are the same and equal

$$\begin{aligned}
&\mathrm{E}\Big[\tau_1^2\bar{\varepsilon}_k^2 - 2\tau_1\bar{\tau}\bar{\varepsilon}_k^2 + \bar{\tau}^2\bar{\varepsilon}_k^2 - 2\tau_1^2\bar{\varepsilon}_k\bar{\bar{\varepsilon}} + 4\tau_1\bar{\tau}\bar{\varepsilon}_k\bar{\bar{\varepsilon}} - 2\bar{\tau}^2\bar{\varepsilon}_k\bar{\bar{\varepsilon}} + \tau_1^2\bar{\bar{\varepsilon}}^2 - 2\tau_1\bar{\tau}\bar{\bar{\varepsilon}}^2 + \bar{\bar{\varepsilon}}^2\bar{\tau}^2\Big] \\
&= \frac{1}{n}\sigma_M^2\sigma_\varepsilon^2 - \frac{2}{Kn}\sigma_M^2\sigma_\varepsilon^2 + \frac{1}{K^2n}\mathrm{E}[\tau^2\varepsilon^2] + \frac{K-1}{K^2n}\sigma_M^2\sigma_\varepsilon^2 - \frac{2}{Kn}\sigma_M^2\sigma_\varepsilon^2 + \frac{4}{K^2n}\sigma_M^2\sigma_\varepsilon^2 \\
&\quad - \frac{2}{K^3n}\mathrm{E}[\tau^2\varepsilon^2] - \frac{2(K-1)}{K^3n}\sigma_M^2\sigma_\varepsilon^2 + \frac{1}{K^2n}\mathrm{E}[\tau^2\varepsilon^2] + \frac{K-1}{K^2n}\sigma_M^2\sigma_\varepsilon^2 \\
&\quad - \frac{2}{K^3n}\mathrm{E}[\tau^2\varepsilon^2] - \frac{2(K-1)}{K^3n}\sigma_M^2\sigma_\varepsilon^2 + \frac{K-1}{K^3n}\sigma_M^2\sigma_\varepsilon^2 + \frac{1}{K^3n}\mathrm{E}\left[\tau^2\varepsilon^2\right] \\
&= \frac{2K-3}{K^3n}\mathrm{E}\left[\tau^2\varepsilon^2\right] + \frac{K^3-2K^2-K+3}{K^3n}\sigma_M^2\sigma_\varepsilon^2.
\end{aligned}$$

Likewise $\mathrm{E}\left[(\tau_1-\bar{\tau})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})^2\right]$ and $\mathrm{E}\left[(\tau_k-\bar{\tau})(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})^2\right]$ are the same and equal

$$\begin{aligned}
&\mathrm{E}\left[(\tau_1\bar{\varepsilon}_1-\bar{\tau}\bar{\varepsilon}_1-\tau_1\bar{\bar{\varepsilon}}+\bar{\tau}\bar{\bar{\varepsilon}})(\bar{\varepsilon}_k^2-2\bar{\varepsilon}_k\bar{\bar{\varepsilon}}+\bar{\bar{\varepsilon}}^2)\right] \\
&\quad = \mathrm{E}\Big[\tau_1\bar{\varepsilon}_1\bar{\varepsilon}_k^2-\bar{\tau}\bar{\varepsilon}_1\bar{\varepsilon}_k^2-\tau_1\bar{\bar{\varepsilon}}\bar{\varepsilon}_k^2+\bar{\tau}\bar{\bar{\varepsilon}}\bar{\varepsilon}_k^2-2\tau_1\bar{\varepsilon}_1\bar{\varepsilon}_k\bar{\bar{\varepsilon}}+2\bar{\tau}\bar{\varepsilon}_1\bar{\varepsilon}_k\bar{\bar{\varepsilon}} \\
&\qquad\qquad +2\tau_1\bar{\varepsilon}_k\bar{\bar{\varepsilon}}^2-2\bar{\tau}\bar{\varepsilon}_k\bar{\bar{\varepsilon}}^2+\tau_1\bar{\varepsilon}_1\bar{\bar{\varepsilon}}^2-\bar{\tau}\bar{\varepsilon}_1\bar{\bar{\varepsilon}}^2-\tau_1\bar{\bar{\varepsilon}}^3+\bar{\tau}\bar{\bar{\varepsilon}}^3\Big] \\
&\quad = \frac{1}{K^2}\mathrm{E}[\tau\bar{\varepsilon}^3]-\frac{2}{K^3}\mathrm{E}[\tau\bar{\varepsilon}^3]+\frac{1}{K^2}\mathrm{E}[\tau\bar{\varepsilon}^3]-\frac{1}{K^3}\mathrm{E}[\tau\bar{\varepsilon}^3]-\frac{1}{K^3}\mathrm{E}[\tau\bar{\varepsilon}^3]+\frac{1}{K^3}\mathrm{E}\left[\tau\bar{\varepsilon}^3\right] \\
&\quad = \frac{2K-3}{K^3n^2}\mathrm{E}\left[\tau\varepsilon^3\right],
\end{aligned}$$

where the second equality follows because τ_1, τ_k, $\bar{\varepsilon}_1$, and $\bar{\varepsilon}_k$ are all uncorrelated and have zero mean. Consequently, after expanding the long expression in square brackets to a weighted sum of terms of the form $\tau_h\bar{\varepsilon}_i\bar{\varepsilon}_j^2$, only terms with $h=i=j$ have nonzero expectation. The last term in Equation (22) is

$$\begin{aligned}
&\mathrm{E}\left[(\tau_1-\bar{\tau})(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}})(\tau_k-\bar{\tau})(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}})\right] \\
&\quad = \mathrm{E}\left[\bar{\tau}\tau_k\bar{\varepsilon}_1\bar{\bar{\varepsilon}}\right]+\mathrm{E}[\bar{\tau}\tau_1\bar{\varepsilon}_k\bar{\bar{\varepsilon}}]+\mathrm{E}[\bar{\tau}\tau_1\bar{\varepsilon}_1\bar{\bar{\varepsilon}}]+\mathrm{E}[\bar{\tau}\tau_k\bar{\varepsilon}_k\bar{\bar{\varepsilon}}] \\
&\qquad -\mathrm{E}[\bar{\tau}^2\bar{\varepsilon}_1\bar{\bar{\varepsilon}}]-\mathrm{E}[\bar{\tau}^2\bar{\varepsilon}_k\bar{\bar{\varepsilon}}]-\mathrm{E}\left[\bar{\tau}\tau_k\bar{\bar{\varepsilon}}^2\right]-\mathrm{E}\left[\bar{\tau}\tau_1\bar{\bar{\varepsilon}}^2\right]+\mathrm{E}\left[\bar{\tau}^2\bar{\bar{\varepsilon}}^2\right] \\
&\quad = \frac{2}{K^2n}\sigma_M^2\sigma_\varepsilon^2+\frac{2}{K^2n}\mathrm{E}\left[\tau^2\varepsilon^2\right]-\frac{2}{K^3}\left(\frac{K-1}{n}\sigma_M^2\sigma_\varepsilon^2+\frac{1}{n}\mathrm{E}\left[\tau^2\varepsilon^2\right]\right) \\
&\qquad\qquad -\frac{2}{K^3}\left(\frac{K-1}{n}\sigma_M^2\sigma_\varepsilon^2+\frac{1}{n}\mathrm{E}\left[\tau^2\varepsilon^2\right]\right)+\frac{(K-1)}{K^3n}\sigma_M^2\sigma_\varepsilon^2+\frac{1}{K^3n}\mathrm{E}\left[\tau^2\varepsilon^2\right] \\
&\quad = \frac{3-2K}{K^3n}\sigma_M^2\sigma_\varepsilon^2+\frac{2K-3}{K^3n}\mathrm{E}\left[\tau^2\varepsilon^2\right].
\end{aligned}$$

Substituting these results into Equation (22), subtracting the square of Equation (18), and multiplying by Kn^2 yields

$$\begin{aligned}
&Kn^2\sum_{k=2}^{K}\mathrm{Cov}\left[((\tau_1-\bar{\tau})+(\bar{\varepsilon}_1-\bar{\bar{\varepsilon}}))^2,((\tau_k-\bar{\tau})+(\bar{\varepsilon}_k-\bar{\bar{\varepsilon}}))^2\right] \\
&= \frac{(2K-3)(K-1)n^2}{K^2}\mathrm{E}[\tau^4]-\frac{(4K-9)(K-1)n^2}{K^2}\sigma_M^4-\frac{(4K-9)(K-1)}{K^2}\sigma_\varepsilon^4 \\
&\quad -\frac{2n(4K-9)(K-1)}{K^2}\sigma_M^4\sigma_\varepsilon^4+\frac{(2K-3)(K-1)}{K^2n}\mathrm{E}[\varepsilon^4]+\frac{3(K-1)(2K-3)(n-1)}{K^2n}\mathrm{E}[V^2] \\
&\quad +\frac{4(K-1)(2K-3)}{K^2}\mathrm{E}[\tau\varepsilon^3]+\frac{6n(K-1)(2K-3)}{K^2}\mathrm{E}[\tau^2\varepsilon^2].
\end{aligned}$$

This expression is the second term in (16); the first term of (16) was computed to be (20).

Summing these two terms, we finally have

$$\begin{aligned}
\mathrm{Var}\left[\mathrm{SS}_\tau\right] \;=\;& \frac{n^2(K-1)^2}{K}\mathrm{E}\left[\tau^4\right] - \frac{(K-1)(K-3)n^2}{K}\sigma_M^4 - \frac{(K-1)(K-3)}{K}\sigma_\varepsilon^4 \\
& - \frac{2(K-1)(K-3)n}{K}\sigma_M^2\sigma_\varepsilon^2 + \frac{(K-1)^2}{Kn}\mathrm{E}\left[\varepsilon^4\right] + \frac{3(K-1)^2(n-1)}{Kn}\mathrm{E}\left[V^2\right] \\
& + \frac{4(K-1)^2}{K}\mathrm{E}\left[\tau\varepsilon^3\right] + \frac{6(K-1)^2 n}{K}\mathrm{E}\left[\tau^2\varepsilon^2\right]. \qquad (23)
\end{aligned}$$

B Pilot Estimation

Suppose that we have the output of a simulation experiment in which there are K_0 scenarios and an inner-level sample size of n_0. To use Equation (12), we need estimates of three quantities: $\mathrm{E}[V^2]$, σ_M^4, and $\mathrm{E}[\tau^4]$. In this appendix, we propose estimators that have some justification, although they are not optimal. For our example of Section 4, we found that these estimates were adequate for the purpose of choosing n^* if K_0 and n_0 were large. We estimate $\mathrm{E}[V^2] = \mathrm{E}[\mathrm{E}[\varepsilon^2|Z]^2]$ by

$$\frac{1}{K_0}\sum_{k=1}^{K_0}\left(\frac{1}{n_0-1}\sum_{j=1}^{n_0} X_{kj}^2 - \frac{(\sum_{j=1}^{n_0} X_{kj})^2}{n_0}\right)^2.$$

A natural estimator of σ_M^4 is

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

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

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

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:

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

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

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

If n_0 is large, we can approximate $\mathrm{E}\left[\sum_{k=1}^{K_0}(\bar{X}_k-\bar{\bar{X}})^4/K_0\right]$ by

$$\frac{(K_0-1)^4+(K_0-1)}{K_0^4}\mathrm{E}[\tau^4]+\frac{3(K_0-1)(2K_0-3)}{K_0^3}\sigma_M^4+\frac{(K_0-1)^4+(K_0-1)}{K_0^4 n_0}\mathrm{E}[\tau^2\varepsilon^2].$$

Solving 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 inner-level sample size n_0.

C The Delta-Hedging Example

The example is of delta-hedging a portfolio that is short one European put option and one European call option on the same underlying stock, with the same strike price Q and maturity 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. P&L is also measured at maturity T. The example is very similar to one used by Baysal et al. (2008, §3), which can be consulted for details not supplied here. Here we focus on formulating the example in a way that fits our ANOVA framework.

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. Thus, the number of shares to hold, $-\Delta$, is the sum of the deltas of the put and the call. 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

$$(p_0+\Delta_0 S_0)e^{rT}+\sum_{i=1}^{s}(\Delta_i-\Delta_{i-1})S_i e^{r(T-t_i)}-|S_s-Q| \tag{24}$$

where p_0 and $|S_s - Q|$ are respectively the initial price and the payoff of the options.

In many models, it is straightforward to simulate the path of stock prices $S_1, \ldots, S_s$, which constitutes a scenario Z in the ANOVA framework, but not easy to compute Δ_i as a function of this path. We next exhibit a random variable X such that the P&L in scenario Z is $M = \mathrm{E}[X|Z]$. This 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. In the context of the nested simulation illustrated in Figure 3, $Z_k = (S_{k1}, \ldots, S_{ks})$,

$$\begin{aligned} \psi_{kji} &= -e^{-r(T-t_i)} \frac{S_{kji}}{S_{ki}} \mathrm{sign}(S_{kji} - Q), \quad \text{and} \\ X_{kj} &= (p_0 + \Delta_0 S_0) e^{rT} + \sum_{i=1}^{s} (\psi_{kji} - \psi_{k,j,i-1}) S_{ki} e^{r(T-t_i)} - |S_{ks} - Q| \end{aligned}$$

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}. The conditional expectation of X_{kj} given Z_k is the P&L given in Equation (24) because of the unbiasedness of the pathwise sensitivity estimation.

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}.