Proceedings of the 2010 Winter Simulation Conference
B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, eds.

AN EFFICIENT SIMULATION PROCEDURE FOR POINT ESTIMATION OF EXPECTED SHORTFALL

Ming Liu
Barry L. Nelson
Jeremy Staum

Department of Industrial Engineering and Management Sciences
Northwestern University
2145 Sheridan Road, Evanston, IL 60208-3119, U.S.A.

ABSTRACT

We present a computationally efficient simulation procedure for point estimation of expected shortfall. The procedure applies tools for ranking and selection to allocate more computational resources to estimation of the largest losses, which are those that affect expected shortfall. Given a fixed computational budget, our procedure estimates expected shortfall with a much lower mean squared error than a standard simulation procedure and much more precisely than an existing interval estimation procedure.

1 INTRODUCTION

Nested simulation makes some risk management problems computationally challenging. To measure risk, we may want to know the value of our portfolio in many possible future scenarios, generated by historical simulation or Monte Carlo sampling from a distribution of relevant risk factors. If our portfolio contains derivative securities, we may need to use Monte Carlo simulation to estimate its value in each scenario. The resulting computational burden can be quite large, with thousands of Monte Carlo replications performed in each of thousands of scenarios, for a total of millions of replications. Researchers have developed two approaches to making nested simulation more computationally efficient.

Authors such as Frye (1998) and Shaw (1998) proposed to use dimension reduction and interpolation to approximate the portfolio value in all scenarios after performing Monte Carlo simulations to estimate the portfolio value in only some scenarios. That is, even though scenarios may involve a high-dimensional vector of risk factors, the analyst identifies just a few principal components that explain most of the variability in portfolio value, chooses certain scenarios in which to value the portfolio by simulation, and then uses linear interpolation among those simulated values to approximate the portfolio value in all other scenarios. This approach is promising, but it requires analyst effort to perform dimension reduction and to design and validate the simulation experiment, since little is known about how to quantify or reduce the errors caused by interpolation and by Monte Carlo sampling.

The other approach is more automated and generic. The earliest work is the thesis of Lee (1998), who studied point estimation of a quantile of the distribution of a conditional expectation. This is related to point estimation of value at risk (VaR): let the portfolio value V in scenario Z be $V(Z) = \mathrm{E}[X|Z]$, where X is the discounted payoff of the securities in the portfolio and E represents risk-neutral expectation. Recently, this literature has focused on two-level simulation, in which the outer level of simulation samples scenarios, and the inner level samples payoffs conditional on the scenarios. Lee (1998) discusses how to reduce the mean squared error (MSE) of the point estimator by jackknifing to reduce its bias and by choosing the number of scenarios to sample in an asymptotically optimal way. Gordy and Juneja (2006, 2008) use similar ideas in proposing a simulation procedure for point estimation of a portfolio's VaR via two-level simulation. Broadie, Du, and Moallemi (2010) develop efficient sequential simulation procedures for estimating the probability that loss exceeds a specified threshold, which is closely related to VaR. Expected shortfall (ES) is another widely used risk measure, closely related to conditional value at risk and tail conditional expectation, which is the conditional expectation of loss given that it exceeds VaR. Gordy and Juneja (2008) mention ES but do not provide a simulation procedure for estimating it. A two-level simulation procedure for interval estimation of ES is the topic of Lan, Nelson, and Staum (2007, 2010), who increase computational efficiency by dynamic allocation of the computational budget in multi-stage simulation.

We focus on point estimation of ES and on the inner level of simulation. Our methods are related to those of Lan, Nelson, and Staum (2010) and of Lesnevski, Nelson, and Staum (2008), who considered another risk measure,

but we apply them differently because our goal is efficient point estimation. To get an estimator with low MSE, we create a heuristic simulation procedure. Although we present some justifications for our heuristics based on the assumption that the simulated data is normally distributed, we do not prove anything about the performance of the procedure. We merely craft and explain a simulation procedure, then use experiments with normal and non-normal data to show that it performs well. More details of our procedure can be found in Liu, Nelson, and Staum (2008) and Liu (2010). Our procedure can attain a sufficiently low MSE even when the computational budget is so small that other methods for estimating ES yield answers that are not accurate enough to be useful. We compare our method to a standard two-level simulation of ES without any efficiency techniques, and to the confidence interval procedure of Lan, Nelson, and Staum (2010). We report experimental results in which our procedure delivers root mean squared error (RMSE) between 1% and 10% of the true ES while the RMSE of a standard two-level simulation and the confidence interval width of Lan, Nelson, and Staum (2010) are about the same magnitude as ES, indicating that those procedures' answers are not useful.

2 EXPECTED SHORTFALL

Let V be a random variable denoting the value of a portfolio at a future time T. Its cumulative distribution function is denoted by F_V. A risk measure, such as VaR or ES, is a functional $T(F_V)$ of this distribution. The expected shortfall at level $1-p$ is defined as

$$\mathrm{ES}_{1-p} = -\frac{1}{p}\left(\mathrm{E}[V\mathbf{1}_{\{V\le v_p\}}] + v_p(p - \Pr[V \le v_p])\right) \tag{1}$$

where v_p is the p-quantile of F_V; $-v_p$ is VaR at the $1-p$ level. In our analysis, we will assume that F_V is continuous at v_p, so that the second term on the right side of Equation (1) vanishes, but our procedure works even if this is not so.

Let us suppose we have k scenarios describing the state of the financial markets at time T. Each scenario specifies the levels of a vector Z of risk factors that determine the portfolio's value V. Examples of risk factors are underlying asset prices, volatilities, or interest rates. Define $V_i = \mathrm{E}[X|Z = Z_i]$, the value of the portfolio in scenario i, expressed as a conditional risk-neutral expectation of the total discounted payoff X of the securities in the portfolio. To simplify notation, we let X_i represent a random variable whose distribution is the conditional distribution of X given $Z = Z_i$, so that $V_i = \mathrm{E}[X_i]$, and we refer to X_i as a "payoff." The expectation is estimated by Monte Carlo simulation.

Let π_V be a permutation of $\{1,2,\ldots,k\}$ such that $V_{\pi_V(1)} \le V_{\pi_V(2)} \le \ldots \le V_{\pi_V(k)}$, that is, scenario $\pi_V(i)$ is the one in which the portfolio value is the ith lowest. Also define γ to be the set of the $\lceil kp \rceil$ portfolios with the smallest values, i.e., $\gamma = \{\pi_V(1), \pi_V(2), \ldots, \pi_V(\lceil kp \rceil)\}$. We use the terms "tail" and "non-tail" to refer to γ and $\{1,2,\ldots,k\} \setminus \gamma$, respectively. Then ES at level $1-p$ of the empirical distribution of $V_1, V_2, \ldots, V_k$ is

$$\mathrm{ES}_{1-p} = \sum_{i=1}^{\lceil kp \rceil} w_i V_{\pi_V(i)} \text{ where } w_i = \begin{cases} -1/kp, & \text{for } i = 1,\ldots,\lfloor kp \rfloor, \\ -1 + \lfloor kp \rfloor / kp, & \text{for } i = \lfloor kp \rfloor + 1. \end{cases} \tag{2}$$

The efficient procedure we propose focuses on estimating ES as specified by Equation (2) when the scenarios are given. The scenarios could be generated by historical data or sampled from a distribution F_V. If we sample them, this represents the outer level of a two-level simulation procedure. We focuse on inner-level simulation, estimating the value of the portfolio in each scenario by simulating payoffs. We will give examples of historical simulation and two-level simulation in §5.

3 THE STANDARD PROCEDURE

In this section, we present the simplest possible simulation procedure for estimating ES as specified by Equation (2). There is a fixed computational budget expressed as a total number C of payoffs that can be simulated. The standard procedure divides the budget equally among the k scenarios and then treats the resulting sample average payoffs as though they were the true values of the scenarios. The procedure is:

1. Simulate payoffs X_{ih} for $i = 1,2,\ldots,k$, $h = 1,2,\ldots,\lfloor C/k \rfloor$. Calculate sample averages $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k$.
2. Select the $\lceil kp \rceil$ smallest sample averages $\bar{X}_{(1)}, \bar{X}_{(2)}, \ldots, \bar{X}_{(\lceil kp \rceil)}$, where the subscript (i) denotes the scenario with the ith smallest sample average.
3. The standard estimator of ES is $\sum_{i=1}^{\lceil kp \rceil} w_i \bar{X}_{(i)}$.

There are two main reasons that this standard procedure does not work well when the budget C is small. First, from the standard estimator, we see that only $\lceil kp \rceil$ sample averages actually appear in the estimator, which means that only about pC payoffs appear. The other $(1-p)C$ payoffs are used solely to eliminate $k - \lceil kp \rceil$ scenarios. This way of selecting $\lceil kp \rceil$ scenarios to use in the estimator is inefficient. The second reason is that for all $i = 1,2,\ldots,\lceil kp \rceil$, $\bar{X}_{(i)}$ is

a biased estimator of $V_{\pi_Y(i)}$, due to selection bias. Selection bias is defined as $\mathrm{E}[\bar{X}_i | i \in \hat{\gamma}] - \mathrm{E}[\bar{X}_i]$, where $\hat{\gamma}$ is the set of scenarios corresponding to the $\lceil k\rho \rceil$ smallest sample averages. When the budget C is small, the standard estimator can be badly biased.

4 AN EFFICIENT PROCEDURE

In this section we propose an efficient simulation procedure to estimate expected shortfall. This procedure overcomes the two disadvantages of the standard procedure mentioned above and can give an accurate point estimator of ES when the budget is small.

To avoid spending too much of the budget on scenarios which can be easily excluded from the tail, we follow Lan, Nelson, and Staum (2010) and Lesnevski, Nelson, and Staum (2008) in using screening. Screening is a method, based on the t-test, that eliminates ("screens out") some scenarios to concentrate computational resources on the scenarios that are most likely to be in the tail γ. We combine the goal of screening in Lan, Nelson, and Staum (2010), to screen out all non-tail scenarios, with the highly efficient screening tactics of Lesnevski, Nelson, and Staum (2008), that use multiple stages of screening that terminate when a stopping rule judges that screening is no longer a good use of computational resources. That is, at each stage of the simulation procedure, we simulate more payoffs conditional on all surviving scenarios (the scenarios that we have not screened out yet) and screen out more scenarios that now seem unlikely to be in the tail. Thus, we overcome the first disadvantage of the standard procedure by allocating fewer payoffs to the non-tail scenarios.

We overcome the second disadvantage by avoiding selection bias altogether with a technique called "restarting" (Boesel, Nelson, and Kim 2003): we throw out all the payoffs used in screening. After screening, we select a set $\hat{\gamma}$ of scenarios which we believe belong to the tail, and allocate the remaining computational budget to scenarios in $\hat{\gamma}$. We use only the sample averages of these new payoffs in our ES estimator. Those sample averages were not used in the decision about whether or not to include a scenario in $\hat{\gamma}$, which makes $\mathrm{E}[\bar{X}_i | i \in \hat{\gamma}] = \mathrm{E}[\bar{X}_i]$, and then they have no selection bias. This restarting technique is also used in Lan, Nelson, and Staum (2010) and Lesnevski, Nelson, and Staum (2008). The only source of bias in our procedure comes from the possibility that we may choose $\hat{\gamma}$ incorrectly, i.e., unequal to the true tail γ.

An important difference between our screening procedure and those of Lan, Nelson, and Staum (2010) and Lesnevski, Nelson, and Staum (2008) is that we dynamically select the error level of the t-tests at each stage. Because of their goal of providing a confidence interval with a minimum guaranteed coverage probability, Lan, Nelson, and Staum (2010) and Lesnevski, Nelson, and Staum (2008) were restricted to using a pre-specified, very low error level for the t-tests. Our procedure tends to choose higher error levels, thus screening more aggressively and concentrating more of the computational budget on the scenarios whose sample averages are used in the ES estimator.

4.1 OUTLINE OF THE PROCEDURE

We outline our procedure in this section, and elaborate on some steps in subsequent sections. For clarity, we split the procedure into two phases, Phase I and Phase II. Phase I includes multi-stage screening and selection of $\hat{\gamma}$. Phase II allocates the remaining computational budget to the selected scenarios, simulates more payoffs, and computes the ES estimator. Because Phase I contains multiple stages, we use $j = 0, 1, 2, \ldots$ to index the stages.

The user specifies the computational budget C, the sample size n_0 of the first stage, and the rate R at which the cumulative sample size grows from one stage to the next. The computational budget can be chosen based on the time available for the simulation experiment or on experience with the budget required to attain the desired precision. An experiment in §5.2 illustrates that it is not difficult to choose good values of n_0 and R, and leads to the recommendation of $n_0 = 30$ and $R = 1.2$ for most simulation problems.

Define I_j to be the set of scenarios that survive to the beginning of stage j and N_j to be the cumulative number of payoffs simulated for each scenario in I_j after stage j, so $N_0 = n_0$. Given the sample size growth factor R, $N_j = N_{j-1}R$ for $j \geq 1$. Let $\bar{X}_i(j)$ be the sample average of scenario i after stage j, i.e., $\bar{X}_i(j) = N_j^{-1} \sum_{h=1}^{N_j} X_{ih}$. Let $\pi_j(\cdot)$ be a mapping of $\{1, 2, \ldots, |I_j|\}$ to I_j such that $\bar{X}_{\pi_j(1)}(j) \leq \bar{X}_{\pi_j(2)}(j) \leq \cdots \leq \bar{X}_{\pi_j(|I_j|)}(j)$. That is, for any $i = 1, 2, \ldots, |I_j|$, $\pi_j(i)$ is the scenario with the sample average that is ith smallest after stage j among the scenarios in I_j. Let C_j be the remaining budget at the beginning of stage j, and let J be the index of the last screening stage in Phase I, as determined by the stopping rule. Let α_j be the error level of each t-test at stage j, which we refer to as the error level for screening at stage j. An outline of our procedure follows; for the full details, see Liu (2010).

Initialization. Set $N_0 \leftarrow n_0$, $I_0 \leftarrow \{1, 2, \ldots, k\}$, $C_0 \leftarrow C$, and $j \leftarrow 0$.

Phase I.

1. If $j > 0$, set $n_j \leftarrow N_j - N_{j-1}$. Simulate n_j payoffs for each scenario in I_j using common random numbers (CRN; see Law and Kelton (2000)) to sharpen screening in Step 3. Calculate the remaining budget $C_{j+1} \leftarrow C_j - |I_j| n_j$.

2. Choose the error level for screening, α_j (§4.3).
3. **Screening:** Screen to compute I_{j+1}, the set of scenarios that survive screening after stage j (§4.2).
4. If the stopping rule is not satisfied (§4.5), then set $j \leftarrow j+1$ and go to Step 1.
5. **Selection:** Set $J \leftarrow j$ and $\hat{\gamma} \leftarrow \{\pi_J(1), \pi_J(2) \ldots, \pi_J(\lceil kp \rceil)\}$.

Phase II. Restart, allocate the remaining computational budget to scenarios in $\hat{\gamma}$, and compute the ES estimator (§4.4).

4.2 Screening

In this section we present the screening method given the target error level α_j at stage j; we will show how to choose α_j in §4.3. For all ordered pairs (i,r) in $I_j \times I_j$, we consider a t-test of the hypothesis that $V_i \leq V_r$ at significance level α_j. If this hypothesis is rejected, we say scenario i is "beaten" by scenario r, i.e., i is beaten by r if and only if

$$\bar{X}_i(j) > \bar{X}_r(j) + \frac{t_{1-\alpha_j, N_j-1} S_{ir}(j)}{\sqrt{N_j}}$$

where $t_{1-\alpha_j, N_j-1}$ is the $1-\alpha_j$ quantile of the t-distribution with $N_j - 1$ degrees of freedom,

$$S_{ir}^2(j) = \frac{1}{N_j - 1} \sum_{h=1}^{N_j} \left(X_{ih} - X_{rh} - (\bar{X}_i(j) - \bar{X}_r(j))\right)^2$$

is the sample variance of $X_i - X_r$, and $\bar{X}_i(j)$ is the sample average of $X_{i1}, X_{i2}, \ldots, X_{iN_i}$. Scenarios beaten at least $\lceil kp \rceil$ times are screened out, therefore

$$I_{j+1} = \left\{ i : \sum_{r \in I_j} \mathbf{1}\left\{ \bar{X}_i(j) > \bar{X}_r(j) + \frac{t_{1-\alpha_j, N_j-1} S_{ir}(j)}{\sqrt{N_j}} \right\} < kp, i \in I_j \right\}.$$

The use of the t-test is motivated by the observation that, if N_j is sufficiently large, $(\bar{X}_i(j) - \bar{X}_r(j) - (V_i - V_r))\sqrt{N_j}/S_{ir}$ is approximately Student t distributed (Henderson 2006). For convenience in analysis, we will treat each payoff X_i as though it were normal. The adequacy of this assumption of normality in a closely related procedure was evaluated in Lesnevski, Nelson, and Staum (2008). Furthermore, our procedure neither provides a confidence interval nor guarantees a minimum probability of correctly identifying the tail, so the t-tests here do not need to be valid. We merely use them as a tool for decreasing the MSE of our point estimator given a fixed computational budget.

4.3 Error Level for Screening

The purpose of the stopping rule (§4.5) is to make sure that enough of the computational budget is left for Phase II to accurately estimate the values of the scenarios selected in Phase I, so in choosing the error level α_j for screening at stage j, we only consider how this affects the quality of the set $\hat{\gamma}$ of scenarios that we select in Phase I. In particular, define $\mathrm{CS} := \{\hat{\gamma} = \gamma\}$ to be the event of selecting γ at the end of Phase I. We would like to choose $\alpha_0, \alpha_1, \ldots, \alpha_J$ to maximize $\Pr\{\mathrm{CS}\}$, the probability of correct selection. Unfortunately, this maximization problem is too hard to solve, primarily because we cannot express $\Pr\{\mathrm{CS}\}$ in a useful form to allow it to be optimized over the error levels.

However, the principle behind the existence of an optimal choice of α_j is clear. A small α_j means screening cautiously at stage j, not screening out many scenarios, but having a low probability of mistakenly screening out a scenario that really belongs to the tail γ; a large α_j means screening aggressively, screening out many scenarios, but with a larger probability of mistakenly screening out tail scenarios. If $\alpha_0, \alpha_1, \ldots, \alpha_J$ are too big, we are very likely to make screening mistakes. If we do, some scenarios in γ will not be in I_{J+1}, i.e., will not survive screening, which will prevent us from making a correct selection at the end of Phase I. If $\alpha_0, \alpha_1, \ldots, \alpha_J$ are too small, we will probably not screen out many scenarios before we use up so much of the computational budget that we have to go to Phase II, when we must select exactly $\lceil kp \rceil$ scenarios. If the number $|I_{J+1}|$ of scenarios that survive to this point is too large, each one has a small sample size N_J because the computational budget was depleted after a small number $J+1$ of stages. Then we will be forced to choose among many scenarios on the basis of sample averages that have high variance, because their variances are inversely proportional to N_J. This implies a large probability of making selection mistakes at the end of Phase I. In other words, by being too cautious during screening, we would waste much of the computational budget on scenarios that we should have been bold enough to eliminate. Then we would quickly find ourselves in a situation in which we would be forced to guess, on the basis of inadequate information, the identities of the tail scenarios from among a large set of scenarios. We will attempt to choose a moderate α_j that balances the risks of screening mistakes during Phase I and selection mistakes at the end of Phase I.

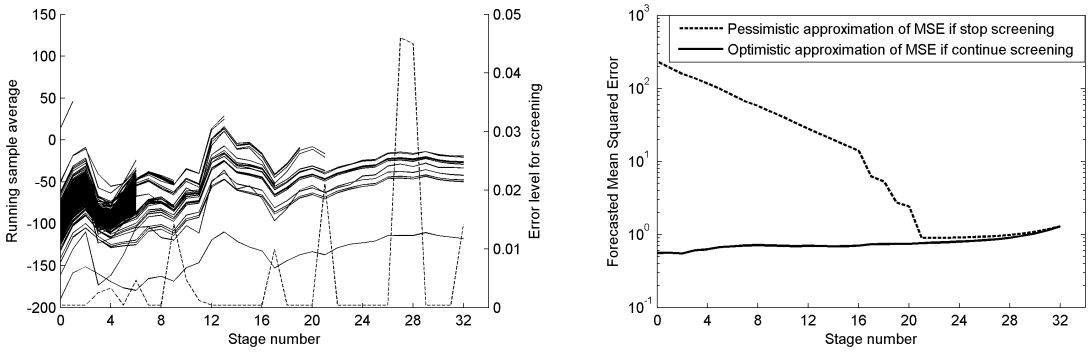

Figure 1: Multi-stage screening (left) and operation of the stopping rule (right) during one run of our procedure on the historical simulation example (§5.2). In the left panel, solid lines represent sample averages of surviving scenarios, and the dashed line is the error level for screening.

Our method chooses $\alpha_0, \alpha_1, \ldots, \alpha_J$ dynamically, on the basis of an approximation to $\Pr\{\mathrm{CS}\}$ that is updated at every stage of screening. We choose the error level α_j at the end of stage j, just before screening. To simplify the problem, we assume while choosing α_j that this error level will be used in screening at the current stage j and all future stages. This is not how our procedure actually works: at stage $j+1$ we will choose α_{j+1} on the basis of new information, and α_{j+1} is generally not the same as α_j. However, the assumption relieves us of the need to consider $\alpha_{j+1}, \alpha_{j+2}, \ldots, \alpha_J$ while choosing α_j, which would be difficult to do.

To choose α_j, we would like to maximize the probability $\Pr\{\mathrm{CS}_j\}$ of selecting all tail scenarios that have survived to stage j: $\mathrm{CS}_j := \{\gamma \cap I_j \subseteq \hat{\gamma}\}$. Unfortunately, we can not write $\Pr\{\mathrm{CS}_j\}$ as an explicit function of α_j; we have to replace it with some sort of approximation. Our approach is to use a forecast of the behavior of our procedure in later stages to construct the following approximation to $\Pr\{\mathrm{CS}_j\}$ when the error level is α:

$$\tilde{P}(j,\alpha) = (1 - \lceil kp \rceil \alpha)^{\tilde{J}(j,\alpha)-j+1} \Big/ \binom{|\tilde{I}(j,\alpha)|}{\lceil kp \rceil}$$

where $\tilde{J}(j,\alpha)$ is the forecasted final stage of Phase I and $\tilde{I}(j,\alpha)$ is the forecasted set of scenarios that will survive screening after stage $\tilde{J}(j,\alpha)$. The procedure for making these forecasts is described in Liu (2010). We choose α_j to maximize $\tilde{P}(j,\alpha)$ instead of $\Pr\{\mathrm{CS}_j\}$, which we can not compute. The derivation of $\tilde{P}(j,\alpha)$ is in Liu (2010). Briefly, the numerator is related to the probability that none of the tail scenarios are screened out in stages $j, j+1, \ldots, \tilde{J}(j,\alpha)$, while the reciprocal of the denominator is related to the probability of correctly choosing $\lceil kp \rceil$ scenarios out of the $|\tilde{I}(j,\alpha)|$ scenarios that are forecasted to survive screening.

The left panel of Figure 1 illustrates how the scenarios' sample averages and the error level for screening α_j change during a single run of the procedure. At many stages, α_j is quite low, because the procedure judges that the number of surviving scenarios is small compared to the remaining computational budget. The same low level is chosen for α_j at many stages because we chose α_j using a search algorithm (Liu 2010) that confines the search to a grid, and this level is the smallest in the grid. At other stages, such as 6, 9, and 21, the procedure judges that there are too many surviving scenarios compared to the remaining budget, so it increases the screening error level α_j and screens out many scenarios. In this run of the procedure, after stage 21, there are only 11 scenarios left, while we must select $kp = 10$. However, even though the 11th scenario is not screened out, the stopping rule takes until stage 32 to decide that screening is no longer worthwhile. This run is atypical; in replications of this example, screening usually stops when only 10 scenarios remain. We chose to present an atypical run because its later stages show that the error level α_j selected by the procedure can vary greatly depending on the remaining budget and the current sample averages, even when the number of surviving scenarios does not change.

4.4 Allocating the Remaining Budget to Compute the Estimator

In this section, we describe Phase II of the procedure. After restarting, it is necessary to allocate the remaining computational budget to scenarios in $\hat{\gamma}$. We do this so as to minimize the variance of the ES estimator. First we describe Phase II simulation and the ES estimator, then derive the optimal allocation.

Conditional on each scenario $i \in \hat{\gamma}$, we simulate M_i payoffs in Phase II and calculate the sample average $\bar{X}_i$. Because we do not do any comparisons between scenarios in Phase II, CRN is not used; typically, independent sampling leads to a lower variance for the ES estimator

$$\widehat{ES}_{1-p} = \sum_{i=1}^{\lceil kp \rceil} w_i \bar{X}_{\pi_J(i)}.$$

Now we consider how to choose the Phase II sample size M_i for $i \in \hat{\gamma}$. Because we use restarting, the bias of the ES estimator only comes from the possibility of a wrong selection $\hat{\gamma} \neq \gamma$ in Phase I. The bias does not depend on Phase II sample sizes, so if we want to minimize the MSE we only need to minimize the variance. The variance of $\widehat{ES}_{1-p}$ is

$$\mathrm{Var}(\widehat{ES}_{1-p}) = \mathrm{Var}\left(\sum_{i=1}^{\lceil kp \rceil} w_i \bar{X}_{\pi_J(i)} \right) = \sum_{i=1}^{\lceil kp \rceil} w_i^2 \frac{\sigma^2_{\pi_J(i)}}{M_{\pi_J(i)}},$$

where $\sigma^2_{\pi_J(i)} = \mathrm{Var}[X_{\pi_J(i)} | \pi_J(i)]$ is the conditional variance of the payoff given that the scenario is $\pi_J(i)$. Notice that, conditional on Phase I, $\mathrm{Var}(\widehat{ES}_{1-p})$ is not a random variable. Since we do not know $\sigma^2_{\pi_J(i)}$, we use the sample variance $S^2_{\pi_J(i)}(J)$ of the N_J samples in Phase I instead. Then we consider the optimization problem

$$\min \sum_{i=1}^{\lceil kp \rceil} w_i^2 \frac{S^2_{\pi_J(i)}(J)}{M_{\pi_J(i)}} \quad \text{s.t.} \quad \sum_{i=1}^{\lceil kp \rceil} M_{\pi_J(i)} = C_{J+1}.$$

Using the Karush-Kuhn-Tucker (KKT) condition, the optimal M_i is

$$M_{\pi_J(i)} = C_{J+1} \frac{w_i S_{\pi_J(i)(J)}}{\sum_{r=1}^{\lceil kp \rceil} w_r S_{\pi_J(r)}(J)}. \tag{3}$$

4.5 Stopping Rule

At the end of each stage in Phase I, our procedure has to decide whether to go on screening, continuing Phase I, or to stop screening, select $\hat{\gamma}$, and end Phase I. Because of restarting, we do not want to continue Phase I too long, or we will throw out a lot of simulated payoffs, leaving too small a computational budget for Phase II, which will produce a high-variance estimator. On the other hand, if we end Phase I too soon, when it is not yet clear which scenarios belong to the tail, a large bias arises because we are likely to select $\hat{\gamma}$ badly. In this section we give a stopping rule for Phase I that balances these considerations.

We focus on the decision whether to stop Phase I after stage j, when I_{j+1} has just been computed. If $|I_{j+1}| = \lceil kp \rceil$, there is no need to do any more screening, so we stop. Otherwise, we approximate the MSE of the ES estimator if we stop now and if we continue, then make the decision that leads to the smallest MSE. In approximating the MSE if we stop now, our procedure is pessimistic about the bias of the ES estimator. In approximating the MSE if we continue, our procedure is optimistic in believing that only scenarios belonging to the tail will survive screening at stage $j+1$, and that these are the scenarios with the smallest conditional payoff variances σ^2 of all scenarios in I_{j+1}. Because we are optimistic about the next stage of screening and pessimistic about stopping, our procedure tends to continue screening when the remaining computational budget C_{j+1} is sufficiently large. As the remaining computational budget shrinks, the variance of the ES estimator grows, and this eventually forces the procedure to stop Phase I to save enough budget for ES estimation in Phase II. We adopt this idea of being pessimistic about stopping and optimistic about continuing because it performed well in Lesnevski, Nelson, and Staum (2008).

Because of restarting, a Phase II sample average $\bar{X}_{\pi_J(i)}$ is an unbiased estimator of $V_{\pi_J(i)}$. Thus the bias of our ES estimator defined in Equation (3) is

$$\mathrm{Bias}(\widehat{ES}_{1-p}) = \mathrm{E}\left[\widehat{ES}_{1-p}\right] - ES_{1-p} = \sum_{i=1}^{\lceil kp \rceil} w_i \left(\mathrm{E}\left[\bar{X}_{\pi_J(i)}\right] - V_{\pi_V(i)} \right) = \sum_{i=1}^{\lceil kp \rceil} w_i \left(V_{\pi_J(i)} - V_{\pi_V(i)} \right). \tag{4}$$

From the definition of w_i, $\pi_V(\cdot)$, and $\pi_j(\cdot)$, it follows that $\mathrm{Bias}(\widehat{ES}_{1-p})$ is negative.

When we consider whether to stop screening after stage j, we can split the bias into two parts: the bias from screening mistakes up to stage j and the bias from any screening or selection mistakes after stage j. The bias due to screening up to stage j is the same whether we stop or continue after stage j, so we ignore it in formulating the stopping rule and only consider the bias due to screening or selection mistakes after stage j. To simplify matters, we

suppose $\gamma \subseteq I_{j+1}$, that is, no screening mistakes have occurred so far. Given our optimistic view of continuing, we suppose there will be no screening or selection mistakes after stage j if we continue, producing zero bias. If we stop after stage j, the only bias comes from selection mistakes due to $|I_{j+1}| > \lceil kp \rceil$: if we select $\hat{\gamma}$ now on the basis of N_j samples from each surviving scenario, it may not be the same as γ. Consistent with our pessimistic approach to evaluating the decision to stop, we consider the following approximate lower bound for the bias (which is negative) due to stopping after stage j:

$$B(j) = \sum_{i=1}^{\min\{\lceil kp \rceil, |I_{j+1}| - \lceil kp \rceil\}} w_i \max_{\delta \geq 0} \delta \Phi\left(-\delta \sqrt{N_j}/\tau_j\right)$$

where $\Phi(\cdot)$ is the standard normal distribution function and $\tau_j = \max\left\{S_{ir}(j) : i, r \in I_{j+1}, i \neq r\right\}$. We refer to Liu (2010) for details of the derivation.

We estimate the variance of the ES estimator if we stop after stage j by

$$V_s(j) = \sum_{i=1}^{\lceil kp \rceil} \frac{w_i^2 S^2_{\pi_j(i)}(j)}{M_{\pi_j(i)}} = \frac{1}{C_{j+1}} \left(\sum_{i=1}^{\lceil kp \rceil} w_i S_{\pi_j(i)}(j) \right)^2.$$

The second equality follows from Equation (3) and $J = j$. Our pessimistic approximation to the MSE of the estimator if we stop after stage j is $MSE_s(j) = B^2(j) + V_s(j)$.

To analyze the variance if we continue, we optimistically suppose that the set of scenarios which will survive after one additional stage of screening is exactly γ, and that they have the smallest variances among the scenarios in I_{j+1}. According to this optimistic assumption, we will stop after stage $j+1$, our ES estimator will have zero bias, and its variance is estimated by

$$V_c(j) = \frac{1}{C_{j+1} - (N_{j+1} - N_j)|I_{j+1}|} \left(\sum_{i=1}^{\lceil kp \rceil} w_i S_{\pi_{S(j)}(i)}(j) \right)^2$$

where $\pi_{S(j)}(\cdot)$ is a mapping of $\left\{1, 2, \ldots, |I_{j+1}|\right\}$ to I_{j+1} such that $S_{\pi_{S(j)}(1)}(j) \leq S_{\pi_{S(j)}(2)}(j) \leq \cdots \leq S_{\pi_{S(j)}(|I_{j+1}|)}(j)$; i.e., $S_{\pi_{S(j)}(i)}(j)$ is the ith smallest sample standard deviation among the scenarios surviving stage j. Our optimistic approximation to the MSE of the estimator if we continue after stage $j+1$ is $MSE_c(j) = V_c(j)$.

The stopping rule is: if $|I_{j+1}| = \lceil kp \rceil$ or $MSE_s(j) < MSE_c(j)$, select $\hat{\gamma}$ and go to Phase II, otherwise continue with stage $j+1$ of screening in Phase I. This rule determines when Phase I ends, but we also use it while choosing the screening error level α_j (§4.3) to forecast when Phase I will end. When we use the stopping rule for that purpose, we plug the forecasted sample averages, sample variances and sets of surviving scenarios into the MSE expressions given above.

The right panel of Figure 1 shows how the stopping rule works on the same run of our procedure shown in Figure 1. The pessimistic approximation of MSE if we stop drops steeply at stages 13, 17, and 21, as the number of surviving scenarios gets close to $kp = 10$. As mentioned previously, this run is atypical in that 11 scenarios survive from stages 21 to 32. On this run, optimism that the sole surviving non-tail scenario will be screened out is not borne out. Both estimates, MSE_c and MSE_s, of MSE rise after stage 21 as the computational budget is spent without achieving anything, but MSE_c rises faster because it includes the effects of continuing for one more extra, larger stage of screening. When it catches up to MSE_s, Phase I ends and the procedure selects the 10 scenarios with the lowest sample averages in the left panel of Figure 1. On this run, 53% of the budget was spent in Phase I.

5 EXPERIMENTAL RESULTS

We test the performance of our procedure on three examples. The first example features artificial configurations of $k = 1000$ scenarios in which the payoffs have heavy-tailed Pareto distributions. We vary a parameter that controls the difficulty of screening and selection and illustrate that our procedure attains lower MSE than the standard procedure (§3) for all values of the parameter that we considered. The second example is of a portfolio of eight call options, with 1000 scenarios based on historical stock prices. The third example is similar to the second, but it is a two-level simulation, with scenarios sampled in an outer-level simulation, and our procedure governing the inner-level simulation. Using this example, we compare our procedure with the standard procedure and with the confidence interval procedure of Lan, Nelson, and Staum (2010).

We compare the precision of the estimators these procedures produce given a computational budget expressed in total payoffs simulated. This comparison is not entirely fair because it excludes the overhead of screening, choosing error levels for screening, etc. Excluding the overhead of our procedure and that of Lan, Nelson, and Staum (2010) is

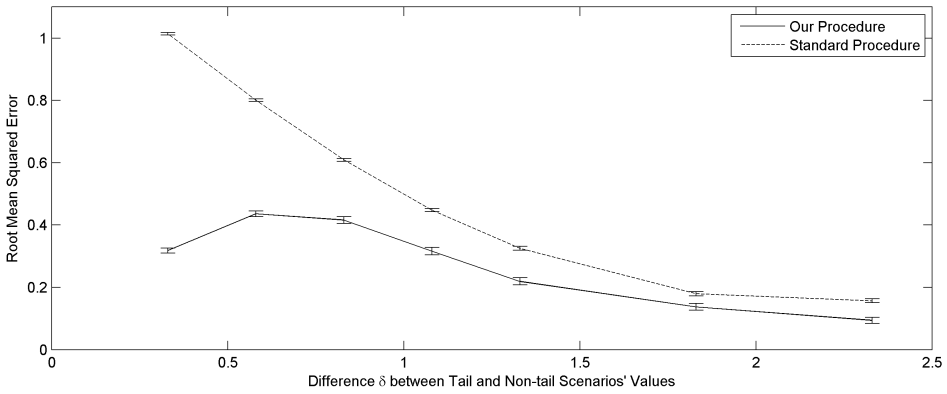

Figure 2: Root mean squared error of estimating expected shortfall at the 99% level, in artificial configurations of varying difficulties, with computational budget $C = 4$ million, first-stage sample size $n_0 = 300$, and sample size growth factor $R = 1.2$.

unfavorable to the standard procedure. This issue is addressed experimentally by Lan, Nelson, and Staum (2010), who also report comparisons in which there is a fixed budget of computing time.

5.1 Artificial Configuration Example

The artificial configuration of scenarios in this example is motivated by the "slippage configuration" used in the ranking and selection literature because it is difficult for screening and selection procedures (Kim and Nelson 2006). In this configuration all tail scenarios have payoffs with a common distribution, while all non-tail scenarios' payoffs have a different common distribution. To make screening even more difficult, the payoffs of different scenarios are independent, so that common random numbers achieve nothing. In particular, if scenarios i and r are both in the tail γ, then $V_i = V_r$; and the $V_i = V_r$ also if neither i nor r are in the tail. If $i \in \gamma$ while $r \notin \gamma$, then $V_r = V_i + \delta$. The parameter δ governs the difficulty of screening and selection: when δ is small, it is difficult to distinguish tail from non-tail scenarios, so it will be hard to screen out scenarios and easy to make selection mistakes. On the other hand, the bias induced by selection mistakes will be small. By changing δ, we can compare our procedure to the standard procedure for a range of configurations with different characteristics.

Pareto distributions are often used to model heavy-tailed loss distributions. Using a heavy-tailed distribution challenges our procedure, which was designed with normally distributed data in mind. We use the Pareto distribution with cumulative distribution function $F(x) = 1 - (\lambda/(\lambda + x))^{2.5}$ for $x \geq 0$. The shape parameter is 2.5 and the scale parameter λ is 25 for tail scenarios, while for non-tail scenarios it is either 25.5, 25.875, 26.25, 26.625, 27, 27.75, or 28.5. The resulting values of δ, the difference between tail and non-tail scenarios' values, are 0.33, 0.58, 0.83, 1.08, 1.33, 1.83, or 2.33. There are $k = 1000$ scenarios and we estimate $\mathrm{ES}_{0.99}$, so there are $kp = 10$ tail scenarios. This example is simple enough that we can compute $\mathrm{ES}_{0.99} = 16.67$, which makes it easier to determine the MSE of the simulation procedures.

Figure 2 shows the root mean squared error (RMSE) of estimating $\mathrm{ES}_{0.99}$ for the standard procedure and our procedure. RMSE was estimated by running 1000 macro-replications of the simulation experiment, and the error bars represent the resulting 95% confidence interval for RMSE. In these experiments, the computational budget C is 4 million payoffs, the initial sample size $n_0 = 300$, and the sample size growth factor $R = 1.2$. From Figure 2 we see that as δ decreases, the RMSE of the standard procedure increases. The reason is that its selection bias increases: when the tail and non-tail scenarios are similar, it is very likely that some of the 990 non-tail scenarios will have sample averages that are less than the value of the tail scenarios and will be selected into $\hat{\gamma}$, the set of scenarios the procedure guesses are in the tail. Because our procedures eliminates selection bias by restarting, it gives a much more accurate point estimator when δ is small. When δ is big, our procedure outperforms the standard procedure because it allocates the computational budget more efficiently. In this experiment, our procedure always yields an RMSE below 0.44, which is small compared to the true $\mathrm{ES}_{0.99} = 16.67$ and to the standard deviation of the tail scenarios' payoff distribution, which is 37.27.

Table 1: Comparison of our procedure with the standard procedure for historical simulation of a portfolio of stock options, with computational budget $C = 4$ million, first-stage sample size $n_0 = 300$, and sample size growth factor $R = 1.2$.

	Method	Variance	Bias	RMSE	SE of RMSE
$ES_{0.99}$	Standard Procedure	23.5	36.7	37.1	0.15
	Our Procedure	0.93	-0.01	0.97	0.02
$ES_{0.95}$	Standard Procedure	5.0	35.4	35.4	0.07
	Our Procedure	2.21	≈ 0	1.49	0.04

5.2 Historical Simulation of an Options Portfolio

Next we consider a more realistic example, in which we estimate the ES of a portfolio of eight call options on Cisco (CSCO) and Sun Microsystems (JAVA). We estimate the ES of this portfolio's value on June 27, 2007 given information up to June 26, 2007: in this example, $T = 1$ day. We use historical simulation, getting $k = 1000$ scenarios from the daily returns on CSCO and JAVA stock, based on their closing prices from July 07, 2003 to June 26, 2007. A scenario consists of the stock prices of CSCO and JAVA on June 27, 2007, created by multiplying their prices on June 26, 2007 (respectively \$27.15 and \$5.01) by one plus their respective returns on a day in the historical data set. We refer to Liu (2010) for details of the simulation model and the options in the portfolio.

Table 1 shows the performance of the standard procedure and our procedure in estimating $ES_{0.99}$ and $ES_{0.95}$. As in the previous example, the computational budget C is 4 million payoffs, the initial sample size $n_0 = 300$, and the sample size growth factor $R = 1.2$. The table also provides the standard error (SE) of estimating each RMSE with 1000 macro-replications. The RMSE of our procedure is significantly smaller than the RMSE of the standard procedure, both in statistical and practical terms. If we define the relative RMSE as the ratio of RMSE to the ES being estimated, we find the relative RMSEs of our procedure for $ES_{0.99}$ and $ES_{0.95}$ are 1.9% and 5.7%, respectively, whereas the standard procedure yields RMSE that is about the same size as ES. Given this budget, our procedure provides moderate accuracy, while the standard procedure provides answers that are not useful and indeed misleading because they are extremely badly biased. It is surprising to see that our procedure delivers a lower RMSE when estimating $ES_{0.99}$ than for $ES_{0.95}$, because it is usually thought to be more difficult to estimate ES deeper in the tail. The primary reason for the surprising result here is that, given this set of 1000 scenarios, it is relatively easy to distinguish the 10 scenarios with the worst losses from the others, but it is not as easy to distinguish the 50 scenarios with the worst losses—for example, the 10th worst loss of \$31.72 is widely separated from the 11th worst loss of \$28.56, but the 50th and 51st worst losses are separated by less than \$0.05, and there are 9 tail scenarios and 13 non-tail scenarios closely packed between the 42nd worst loss of \$16.03 and the 63rd worst loss of \$14.41.

We also tested the sensitivity of our procedure's performance with respect to the first-stage sample size n_0 and the sample size growth factor R, which the user must choose. For estimating $ES_{0.99}$ with computational budget $C = 4$ million, first we fixed $R = 1.2$, and varied n_0. As long as n_0 was between 30 and 1300, RMSE was below 1.11, not far from the best RMSE the procedure attains for any value of n_0. When n_0 was increased past 1300, RMSE increased: it is inefficient to spend a third or more of the computational budget in the first stage, before any scenarios can be screened out. These findings are similar to those of Lesnevski, Nelson, and Staum (2007), and we likewise recommend choosing n_0 to be quite small, but large enough that the first-stage sample averages are approximately normal. Usually, $n_0 = 30$ is large enough (Lesnevski, Nelson, and Staum 2008). Next we fixed $n_0 = 300$ and changed the growth factor R from 1.1 to 2.0. As in Lesnevski, Nelson, and Staum (2007), this had little effect on the procedure's RMSE, which stayed between 0.92 and 1.11. In conclusion, we recommend $R = 1.2$ and $n_0 = 30$, unless the payoff distributions are heavy-tailed (such as the current example), in which case n_0 should be increased until the first-stage sample averages are approximately normal.

5.3 Two-Level Simulation of an Options Portfolio

This example is the same as the example in §5.2, but scenarios are generated differently. Instead of using a fixed set of scenarios drawn from historical data, we generate them in an outer-level simulation. The outer-level simulation samples scenarios from a joint distribution of the two stocks' prices whose parameters are estimated from the historical data. Given the scenarios sampled, the rest of our simulation, i.e., the inner-level simulation, is the same as in §5.2. For a complete description of this example, see Lan, Nelson, and Staum (2010). The purpose of considering this two-level simulation variant of the previous example is to compare our procedure with the two-level simulation procedure of Lan, Nelson, and Staum (2010), which we refer to as the CI procedure because it generates a confidence interval for

Table 2: Comparison of procedures for estimating expected shortfall at the 99% level in a two-level simulation of a portfolio of stock options, with $k = 4000$ scenarios.

Budget	n_0	CI Procedure		Standard Procedure		Our Procedure	
		Average CI Half-width	Standard Error	RMSE	Standard Error	RMSE	Standard Error
4 million	612	164	1.0	109	0.40	6.7	1.6
8 million	1217	104	1.6	69	0.30	1.4	0.11
16 million	2557	49	2.3	41	0.23	0.9	0.07

ES. We make comparisons by sampling scenarios with the CI procedure, then giving these scenarios to the standard procedure or our procedure, which perform inner-level simulation.

In Table 2, we report the average half-width of the 90% confidence interval for $\mathrm{ES}_{0.99}$ generated by the CI procedure and compare it to the RMSEs of the standard procedure and our procedure, using the results of 100 macro-replications. Each procedure uses $k = 4000$ scenarios sampled from the bivariate normal distribution described above. The parameters $k = 4000$ and n_0 listed in the table were chosen by a pilot experiment described in Chapter 4 of Lan (2010) to make the CI procedure perform well. When using our procedure, we set sample size growth factor $R = 1.2$ and used the same n_0 as for the CI procedure, even though it is larger than the best n_0 for our procedure. A large n_0 is good for the CI procedure because it is a two-stage procedure, whereas our multi-stage procedure does well with small n_0. The choice of n_0 is intended to be favorable to the CI procedure and to show that the advantage of our procedure does not depend on picking the best values of the procedure's parameters. The RMSE of our procedure is not exactly comparable to the half-width of a confidence interval, but Table 2 shows our procedure's RMSE is so much smaller than the half-width of the CI procedure that we can conclude that our procedure is greatly preferable given a small computation budget. For these computational budgets, the CI procedure yields a confidence interval whose width is much greater than the ES we are trying to estimate, which is not useful; likewise, the RMSE of the standard procedure is large compared to ES. Our procedure attains a relative RMSE of only a few percent when the budget is 8 or 16 million.

ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Grant No. DMI-0555485. The authors thank Hai Lan for providing assistance with computer code and experiments.

REFERENCES

Boesel, J., B. L. Nelson, and S. Kim. 2003. Using ranking and selection to 'clean up' after simulation optimization. *Operations Research* 51:814–825.

Broadie, M., Y. Du, and C. C. Moallemi. 2010. Efficient risk estimation via nested sequential simulation. Working paper, Columbia University.

Frye, J. 1998. Monte Carlo by day. *Risk* 11:66–71.

Gordy, M. B., and S. Juneja. 2006. Efficient simulation for risk measurement in portfolio of CDOs. In *Proceeedings of the 2006 Winter Simulation Conference*, ed. L. F. Perrone, B. Lawson, J. Liu, and F. P. Wieland, 749–756: IEEE Press.

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

Henderson, S. G. 2006. Mathematics for simulation. In *Simulation*, ed. S. G. Henderson and B. L. Nelson, Volume 13 of *Handbooks in Operations Research and Management Science*, Chapter 2. Amsterdam: Elsevier.

Kim, S.-H., and B. L. Nelson. 2006. Selecting the best system. In *Simulation*, ed. S. G. Henderson and B. L. Nelson, Volume 13 of *Handbooks in Operations Research and Management Science*, Chapter 17. Amsterdam: Elsevier.

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

Lan, H., B. L. Nelson, and J. Staum. 2007. A confidence interval for tail conditional expectation via two-level simulation. In *Proceedings of the 2007 Winter Simulation Conference*, ed. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 949–957: IEEE Press.

Lan, H., B. L. Nelson, and J. Staum. 2010. Confidence interval procedures for expected shortfall via two-level simulation. *Operations Research*. In press.

Law, A. M., and W. D. Kelton. 2000. *Simulation modeling and analysis*. 3rd ed. New York: McGraw-Hill.

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

Lesnevski, V., B. L. Nelson, and J. Staum. 2007. Simulation of coherent risk measures based on generalized scenarios. *Management Science* 53:1756–1769.

Lesnevski, V., B. L. Nelson, and J. Staum. 2008. An adaptive procedure for simulating coherent risk measures based on generalized scenarios. *Journal of Computational Finance* 11:1–31.

Liu, M. 2010. *Efficient simulation in financial risk management*. Ph. D. thesis, Northwestern University.

Liu, M., B. L. Nelson, and J. Staum. 2008. An efficient simulation procedure for point estimation of expected shortfall. Working paper 08-03, Dept. of IEMS, Northwestern University.

Shaw, J. 1998. Beyond VAR and stress testing. In *Monte Carlo: Methodologies and Applications for Pricing and Risk Management*, ed. B. Dupire, 231–244. London: Risk Books.

AUTHOR BIOGRAPHIES

MING LIU is a Ph.D. student in the Department of Industrial Engineering and Management Sciences, McCormick School of Engineering at Northwestern University. His e-mail is ⟨mingliu2010@u.northwestern.edu⟩ and webpage addresses is ⟨users.iems.northwestern.edu/~mingl⟩.

BARRY L. NELSON is the Charles Deering McCormick Professor and Chair of the Department of Industrial Engineering and Management Sciences at Northwestern University. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems. His e-mail and web addresses are ⟨nelsonb@northwestern.edu⟩ and ⟨www.iems.northwestern.edu/~nelsonb/⟩.

JEREMY STAUM is Associate Professor of Industrial Engineering and Management Sciences and holds the Pentair-Nugent Chair at Northwestern University. He coordinated the Risk Analysis track of the 2007 Winter Simulation Conference and serves as department editor for financial engineering at *IIE Transactions* and as an associate editor at *Operations Research*. His website is ⟨users.iems.northwestern.edu/~staum⟩.