Department of Industrial Engineering and Management Sciences

Northwestern University, Evanston, Illinois 60208-3119, U.S.A.

Working Paper No. 08-02

A Confidence Interval Procedure for Expected Shortfall Risk Measurement via Two-Level Simulation

Hai Lan

Barry L. Nelson

Jeremy Staum

Original Draft 20 November 2008

This paper is based upon work supported by the National Science Foundation under Grant No. DMI-0555485.

ABSTRACT

We develop and evaluate a two-level simulation procedure that produces a confidence interval for expected shortfall, a risk measure also known as conditional value-at-risk and worst conditional expectation, and closely related to tail conditional expectation. The outer level of simulation generates financial scenarios while the inner level estimates expected loss conditional on each scenario. Our procedure uses the statistical theory of empirical likelihood to construct a confidence interval, and tools from the ranking-and-selection literature to make the simulation efficient.

1 Introduction

Financial risk management is vital to the survival of financial institutions and the stability of the financial system. A fundamental task in risk management is to measure the risk entailed by a decision, such as the choice of a portfolio. In particular, regulation requires each financial institution to measure the risk of the firm's entire portfolio and to set its capital reserves accordingly, to reduce the chance of bankruptcy if large losses occur. This firm-wide risk measurement problem is challenging. Solution methods that avoid Monte Carlo simulation involve simplifications and approximations that cast doubt on the validity of the answer. Monte Carlo simulation allows for detailed modeling of the behavior of the firm's portfolio given possible future events, and it is compatible with the use of the best available models of financial markets. Because of this, Monte Carlo simulation is an attractive methodology, but its appeal is limited by its computational cost, which can be quite large, especially when derivative securities are involved (McNeil et al., 2005, § 2.3.3). This is because a precise estimate of the risk measure requires consideration of many future financial scenarios, but it takes a long time to compute the value of all the derivative securities in any scenario. Consequently, a large firm-wide risk measurement simulation can take days to run on a cluster of one thousand computers. Because of the speed at which markets move, timelier answers are needed. One of our main contributions is to develop a more efficient simulation procedure for risk measurement when it is time-consuming to compute the portfolio value in a future financial scenario.

Let V be a random variable representing the value of a portfolio in the future, and let F_V be its distribution. A risk measure is a functional $T(F_V)$ of this distribution. For example, value at risk VaR_{1-p} may be defined as the negative of the p-quantile of F_V. In market risk management, it is usual to consider the 95th or 99th percentile: $p = 5\%$ or 1%. In this paper, we focus on expected shortfall:

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

where v_p is the lower p-quantile of F_V. If F_V is continuous at v_p, then ES equals tail conditional expectation (Acerbi and Tasche, 2002):

$$\mathrm{TCE}_{1-p} = -\mathrm{E}[V|V \le v_p].$$

Closed-form expressions for ES are available when the distribution F_V belongs to some simple parametric families (McNeil et al., 2005, §§ 2.2.4, 7.2.3). There is also a literature on nonparametric estimation of expected shortfall from data $V_1, \ldots, V_k$ drawn from a stationary process whose marginal distributions are F_V. In this setting, Chen (2008) shows that although kernel smoothing is valuable in estimating VaR, the simplest nonparametric estimator of ES, involving an average of the $\lceil kp \rceil$ smallest values among $V_1, \ldots, V_k$, is preferred to kernel smoothing. Accordingly, we use unsmoothed averages in our construction of confidence intervals for ES.

However, we consider a different situation, in which we do not have a sample of data from F_V and we do not have a parametric form for F_V. It many risk measurement applications, it is important to consider risk as depending on the current state of the market. In this case, historical loss data is not directly representative of the risks faced today. In particular, suppose that V is the gain experienced by a portfolio containing derivative securities. We have a model of underlying financial markets that allows us to sample a scenario Z (which specifies such things as tomorrow's stock prices) from its distribution F_Z, and there is a function $V(\cdot)$ such that the portfolio's gain $V = V(Z)$. Even if F_Z belongs to a simple parametric family, F_V may not, because the value function $V(\cdot)$ is complicated. Furthermore, if either the model or the derivative securities are complicated, the function $V(\cdot)$ itself is not known in closed form. However, in most models, we can represent $V(Z) = \mathrm{E}[X|Z]$ where X involves the payoffs of derivative securities, which we can simulate conditional on the scenario Z.

In this situation, we can estimate the risk measure $T(F_V)$ by a two-level simulation, in which the outer level of simulation generates scenarios $Z_1, Z_2, \ldots, Z_k$ and the inner level estimates each $V_i := V(Z_i)$ by simulating V conditional on Z_i. For more on this general framework and its significance in risk management, see Lan et al. (2007b). Point estimation of a quantile of the distribution (here called F_V) of a conditional expectation via two-level simulation has been studied by Lee (1998) and Gordy and Juneja (2006, 2008). This is equivalent to point estimation of VaR. Gordy and Juneja (2008) also consider point estimation of ES. This strand of the research literature emphasizes asymptotic optimality for large computational budgets or portfolios. In related work, Steckley and Henderson (2003) study estimation of the density of F_V via two-level simulation.

The present paper focuses on interval estimation of ES and moderate sample sizes, and it improves upon our earlier work in Lan et al. (2007a). We develop a procedure for efficient computation of a confidence interval for ES and show that it performs well at realistic sample sizes. Two-level simulations can be extremely computationally expensive: given the available computational budget, they may produce very wide confidence intervals. To produce a narrower confidence interval given a fixed computational budget, our procedures use screening with common random numbers and allocate sample sizes proportional to each scenario's sample variance. In Appendix A we prove that the coverage of our procedure's confidence interval is at least the nominal level asymptotically. To the best of our knowledge, this paper and our conference papers (Lan et al., 2007a,b) provide the first proof of the asymptotic validity of a confidence interval produced by a two-level simulation; this is one of our main contributions.

Section 2 contains two examples of two-level risk measurement simulation problems. We present our simulation procedure in Section 3. Numerical results of computer experiments in which we apply our procedures to the examples appear in Section 4, while Section 5 concludes and describes future research.

2 Motivating Examples

Risk management simulations may deal with non-trivial models and thousands of complicated securities. However, for purposes of illustration, we consider the following two simple, small examples. This allows us to report the coverage rate that our procedure achieves by repeating the simulation experiment many times, so as to see how often our confidence interval contains the true value of ES.

2.1 Selling a Single Put Option

At time 0, we sell a put option with strike price $K = \$110$ and maturity $U = 1$ year on a stock whose initial price is $S_0 = \$100$. This stock's price obeys the Black-Scholes model with drift $\mu = 6\%$ and volatility $\sigma = 15\%$. There is a money market account with interest rate $r = 6\%$. The initial price for which we sell the put option is $P_0 = P(U, S_0)$, which is the Black-Scholes formula evaluated for maturity U and stock price S_0.

We are interested in $\mathrm{ES}_{0.99}$ at time $T = 1/52$ years, or one week from now. The scenario Z is a standard normal random variable that determines the stock price at time T:

$$S_T = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) T + \sigma\sqrt{T} Z\right).$$

The net payoff at maturity U, discounted to time T, from selling the put for an initial price of P_0 is

$$X = e^{-r(U-T)}\left(P_0 e^{rU} - (K - S_U)^+\right),$$

where

$$S_U = S_T \exp\left(\left(r - \frac{\sigma^2}{2}\right)(U - T) + \sigma\sqrt{U-T}Z'\right)$$

and Z' is a standard normal random variable independent of Z.

In this simple example, we can actually find the value

$$V = \mathrm{E}[X|Z] = P_0 e^{rT} - P(U - T, S_T),$$

using the Black-Scholes formula evaluated for maturity $U - T$ and stock price S_T. Furthermore, V is strictly decreasing in Z, so we can compute that $\mathrm{VaR}_{0.99} \approx \$2.92$ by evaluating V at $Z = z_{0.01}$, the standard normal first percentile. By numerical integration, we can also compute $\mathrm{ES}_{0.99} \approx \$3.39$, which will help us to evaluate the performance of our procedure. (Our procedure does not compute V by using the Black-Scholes formula, but rather estimates it using inner-level simulation of payoffs at maturity.)

2.2 A Portfolio of Options on Two Stocks

We are interested in ES at time $T = 1/365$ years, or one day, of a portfolio of call options on Cisco (CSCO) and Sun Microsystems (JAVA), as shown in Table 1. In the table, the

Table 1: Portfolio of Call Options

Index i	Underlying Stock	Position θ_i	Strike K_i	Maturity U_i	Option Price	Implied Volatility σ_i
1	CSCO	200	\$27.5	0.315	\$1.65	26.66%
2	CSCO	-400	\$30	0.315	\$0.7	25.64%
3	CSCO	200	\$27.5	0.564	\$2.5	28.36%
4	CSCO	-200	\$30	0.564	\$1.4	26.91%
5	JAVA	600	\$5	0.315	\$0.435	35.19%
6	JAVA	1200	\$6	0.315	\$0.125	35.67%
7	JAVA	-900	\$5	0.564	\$0.615	36.42%
8	JAVA	-300	\$6	0.564	\$0.26	35.94%

position given for each option is the number of shares of stock we have the option to buy; if it is negative, then our portfolio is short call options on that many shares of stock. Except for the portfolio weights, which we made up, the data in the table was drawn from listed options prices on June 26, 2007. We ignored the distinction between American and European options because neither of these stocks pays dividends, a situation in which early exercise of an American call option is widely regarded as mistaken (see, e.g., Luenberger, 1998, § 12.4).

The scenario $Z = (Z^{(1)}, Z^{(2)})$ is a bivariate normal random variable that determines the stock prices at time T:

$$S_T^{(j)} = S_0^{(j)} \exp\left(\left(\mu^{(j)} - \frac{1}{2}(\sigma^{(j)})^2\right) T + \sigma^{(j)}\sqrt{T} Z^{(j)}\right), j = 1, 2.$$

Based on sample moments of 1,000 daily stock prices, the volatilities of CSCO and JAVA are respectively $\sigma^{(1)} = 32.85\%$ and $\sigma^{(2)} = 47.75\%$, while the correlation between the components of Z is 0.382. In practice, more sophisticated methods of volatility forecasting would be used, but this method yields a reasonable covariance matrix for the vector S_T of stock prices tomorrow. Because one day is such a short period of time that the effect of the drift μ is negligible, while mean returns are hard to estimate because of the high ratio of volatility to mean, we take each $\mu^{(j)} = 0$.

In addition to a distribution F_Z for scenarios, we need to specify the function $V(\cdot)$ by saying how option values at time T depend on the scenario. We adopt the "sticky strike" assumption, according to which each option's value at time T is given by the Black-Scholes formula with volatility equal to the implied volatility that this option had at time 0 (Derman, 1999). This does not make for an arbitrage-free model of the underlying stock prices S, but it is an assumption that has been used in practice to model short-term changes in option values. As in the previous example, we can compute these values without using inner-level simulation, but our procedure performs inner-level simulation for each option i by taking the stock price at maturity U_i to be

$$S_i = \frac{S_T^{(j_i)}}{D_i} \exp\left(-\frac{1}{2}\sigma_i^2(U_i - T) + \sigma_i\sqrt{U_i - T} Z_i'\right)$$

where $j_1 = j_2 = j_3 = j_4 = 1$ (the four options on CSCO) and $j_5 = j_6 = j_7 = j_8 = 2$ (the four options on JAVA), D_i is a discount factor from T to U_i, and Z' is a standard multivariate normal random vector independent of Z. Based on Treasury bond yields, the discount factor was 0.985 for options maturing in 0.315 years and 0.972 for options maturing in 0.564 years. The independence of the components of Z' means that, even though in reality the eight options depend on two correlated stock prices at two times, independent inner-level simulations are used to estimate the option prices at time T. As shown by Gordy and Juneja (2006, 2008), this can improve the efficiency of the two-level simulation. Furthermore, it is natural to simulate independently for each option under the sticky strike assumption, which does not contain a consistent model of the underlying stock prices. The value of option i at time T is the conditional expectation of the discounted payoff $Y_i := D_i(S_i - K_i)^+$ given $S_T^{(j_i)}$. The profit from holding the portfolio from 0 to T is $V(Z) = \mathrm{E}[X|Z]$ where $X = \theta^\top(Y - P_0/D_0)$ and the discount factor $D_0 \approx 1$ because the time value of money over one day is negligible. We estimated the true value of $\mathrm{ES}_{0.99}$ to be \$32.4, the average point estimate produced by 100 repetitions of the complete experiment with a budget of 1.56 billion inner-level simulations each.

3 The Procedure

This section presents a fixed-budget two-level simulation procedure for generating a confidence interval for ES_{1-p}. The procedure first simulates scenarios $Z_1, Z_2, \ldots, Z_k$. If the values $V_1, V_2, \ldots, V_k$ of these scenarios were known, then the point estimate of ES_{1-p} would be

$$-\frac{1}{p}\left(\sum_{i=1}^{\lfloor kp\rfloor}\frac{1}{k}V_{\pi_V(i)} + \left(p - \frac{\lfloor kp\rfloor}{k}\right)V_{\pi_V(\lceil kp\rceil)}\right) \tag{2}$$

where π_V is a permutation of $\{1, 2, \ldots, k\}$ such that $V_{\pi_V(1)} \leq V_{\pi_V(2)} \leq \cdots \leq V_{\pi_V(k)}$. That is, $V_{\pi_V(i)}$ is the ith order statistic of $V_1, V_2, \ldots, V_k$.

Because these values are not known, they are estimated by inner-level simulation. The inner level of simulation has a first stage in which $n_0 \geq 2$ payoffs are generated for every scenario, using common random numbers (CRN; see, e.g., Law and Kelton, 2000). Let X_i be a random variable representing a payoff simulated under scenario i, that is, having the distribution of the payoff X given $Z = Z_i$. The first-stage sample average of the n_0 payoffs $X_{i1}, X_{i2}, \ldots, X_{in_0}$ is denoted $\bar{X}_i(n_0)$.

After the first stage, screening eliminates scenarios whose values are not likely to appear in Equation (2). Then sample sizes $N_1, N_2, \ldots, N_k$ are chosen; the sample size N_i is 0 if scenario Z_i has been screened out. The goal of screening is to allocate more of the computational budget to scenarios that matter. Next, the first-stage data are discarded, a process called "restarting." This is necessary for the statistical validity of the confidence interval (Boesel et al., 2003).

In the second stage, N_i payoffs $X_{i1}, X_{i2}, \ldots, X_{iN_i}$ are generated conditional on the scenario Z_i for each $i = 1, 2, \ldots, k$ using independent sampling (no CRN). The sample average of

$X_{i1}, X_{i2}, \ldots, X_{iN_i}$ is denoted $\bar{X}_i(N_i)$. Then a confidence interval is formed: the confidence limits appear in Equations (8) and (9) below. The two-level simulation point estimate of ES_{1-p} is

$$-\frac{1}{p}\left(\sum_{i=1}^{\lfloor kp \rfloor} \frac{1}{k} \bar{X}_{\pi_1(i)}(N_{\pi_1(i)}) + \left(p - \frac{\lfloor kp \rfloor}{k}\right) \bar{X}_{\pi_1(\lceil kp \rceil)}(N_{\pi_1(i)})\right) \tag{3}$$

where π_1 is a permutation of $\{1, 2, \ldots, k\}$ such that $\bar{X}_{\pi_1(1)}(N_{\pi_1(1)}) \leq \bar{X}_{\pi_1(2)}(N_{\pi_1(2)}) \leq \cdots \leq \bar{X}_{\pi_1(k)}(N_{\pi_1(i)})$. If $N_i = 0$, then $\bar{X}_i$ is taken to be ∞ so that it is not among the order statistics used in Equation (3).

To get a confidence interval, we need a way of combining uncertainty that arises at the outer level, because $Z_1, Z_2, \ldots, Z_k$ is a sample from F_Z, with uncertainty that arises at the inner level because we only possess an estimate $\bar{X}_i$ of each scenario's value $V_i = V(Z_i)$. In Lan et al. (2007b), we described a framework for two-level simulation that generates a two-sided confidence interval $[\hat{L}, \hat{U}]$ with confidence level $1 - \alpha$ where α can be decomposed as $\alpha = \alpha_{\mathrm{o}} + \alpha_{\mathrm{i}}$, representing errors due to the outer and inner levels of simulation, respectively. Here we further decompose $\alpha_{\mathrm{i}} = \alpha_{\mathrm{s}} + \alpha_{\mathrm{hi}} + \alpha_{\mathrm{lo}}$, where α_{s} is error due to screening and α_{hi} and α_{lo} are errors respectively associated with upper and lower confidence limits for inner-level simulation.

3.1 Screening

Screening is the process of eliminating ("screening out") scenarios to increase the simulation's efficiency by devoting more computational resources to the remaining scenarios. From Equation (2), we can see that ES depends on the values of scenarios $\pi_V(1), \pi_V(2), \ldots, \pi_V(\lceil kp \rceil)$ alone, so we want screening to keep these scenarios but eliminate as many others as possible. Call the set of scenarios that survive screening I, and define $\gamma := \{\pi_V(1), \pi_V(2), \ldots, \pi_V(\lceil kp \rceil)\}$, the set of scenarios we wish to keep. The event of correct screening is $\{\gamma \subseteq I\}$, and we must create a screening procedure such that $\Pr\{\gamma \subseteq I\} \geq 1 - \alpha_{\mathrm{s}}$. The number of pairwise comparisons between γ and all other scenarios is $(k - \lceil kp \rceil)\lceil kp \rceil$. Therefore, for each ordered pair (i, j) we consider a hypothesis test that $V_i \leq V_j$ at level $\alpha_s/((k - \lceil kp \rceil)\lceil kp \rceil)$. If the hypothesis is rejected, then we say Z_i is "beaten" by Z_j. For each $i = 1, 2, \ldots, k$, let $X_{i1}, X_{i2}, \ldots, X_{in_0}$ be an i.i.d. sample drawn from the conditional distribution of X given Z_i, and let $\bar{X}_i(n_0)$ be its sample average. For each $i, j = 1, 2, \ldots, k$, let $S^2_{ij}(n_0)$ be the sample variance of $X_{i1} - X_{j1}, X_{i2} - X_{j2}, \ldots, X_{in_0} - X_{jn_0}$. We retain all risk factors that are beaten fewer than $\lceil kp \rceil$ times:

$$I = \left\{ i : \sum_{i \neq j} \mathbf{1}\left\{ \bar{X}_i(n_0) > \bar{X}_j(n_0) + d\frac{S_{ij}(n_0)}{\sqrt{n_0}} \right\} < \lceil kp \rceil \right\} \tag{4}$$

where $\mathbf{1}\{\cdot\}$ is an indicator function and

$$d = t_{n_0 - 1, 1 - \alpha_s/((k - \lceil kp \rceil)\lceil kp \rceil)} \tag{5}$$

is the $1 - \alpha_s/((k - \lceil kp \rceil)\lceil kp \rceil)$ quantile of the t-distribution with $n_0 - 1$ degrees of freedom.

From Equation (4) we see that it is easier to screen out scenarios when the sample variances $S_{ij}^2(n_0)$ are smaller. We use CRN to reduce the variance of $X_i - X_j$; in financial examples, CRN usually induces a substantial positive correlation between X_i and X_j.

Computing I could require as many as $k(k-1)/2$ comparisons and sample variance computations. However, it is generally unnecessary to do all of them: it is possible to screen out many risk factors with only one comparison each by using a looser but faster heuristic screening test. Before doing screening, we do a "pre-screening" in the following way. Define $\tilde{S}^2(n_0) := \max\{S_{\pi_0(i)}^2 : i = 1, 2, \ldots, \lceil kp \rceil\}$. If $\bar{X}_{\pi_0(i)}(n_0) > \bar{X}_{\pi_0(\lceil kp \rceil)}(n_0) + d\sqrt{(S_{\pi_0(i)}^2(n_0) + \tilde{S}^2(n_0))/n_0}$, and the sample correlations between $X_{\pi_0(i)}$ and $X_{\pi_0(1)}$, $X_{\pi_0(2)}$, $\ldots$, $X_{\pi_0(\lceil kp \rceil)}$ induced by CRN are nonnegative, then scenario i does not belong to I. If i can be screened out in this way, by comparing it only to the $\lceil kp \rceil$th largest sample average, then we can avoid many comparisons and sample covariance computations involving i that appear in Equation (4). The method is heuristic; the proof of the CI's validity in Appendix A applies to the version of the procedure without pre-screening.

3.2 Empirical Likelihood

The procedure uses empirical likelihood (Owen, 2001) to account for statistical uncertainty at the outer level, that is, for the fact that $V_1, V_2, \ldots, V_k$ is only a sample from the true distribution F_V of portfolio values at time horizon T. The construction of an outer-level confidence interval for ES_{1-p} based on empirical likelihood is discussed by Baysal and Staum (2007). Here we review a few essential facts for understanding the operation of empirical likelihood in our two-level simulation procedure.

Empirical likelihood involves assigning a vector $\mathbf{w}$ of weights to the scenarios $Z_1, Z_2, \ldots, Z_k$, or, equivalently, to their values $V_1, V_2, \ldots, V_k$. This weight vector $\mathbf{w}$ must belong to the set

$$\mathcal{S}(k) := \bigcup_{\ell=1}^{k} \mathcal{S}_\ell(k) \text{ where } \mathcal{S}_\ell(k) := \left\{ \mathbf{w} : \mathbf{w} \geq 0,\ \sum_{i=1}^{k} w_i = 1,\ \sum_{i=1}^{\ell} w_i = p,\ \prod_{i=1}^{k} w_i \geq c\, k^{-k} \right\}, \quad (6)$$

where c is a critical value derived from a chi-squared distribution. Each $\mathbf{w} \in \mathcal{S}(k)$ belongs to $\mathcal{S}_\ell(k)$ for a unique integer ℓ. There are integers $\ell_{\min}$ and $\ell_{\max}$ such that $\mathcal{S}_\ell(k)$ is empty if $\ell < \ell_{\min}$ or $\ell > \ell_{\max}$; we need only consider a limited range of ℓ, not all $1, 2, \ldots, k$. Although ℓ depends on $\mathbf{w}$, while $\ell_{\min}$ and $\ell_{\max}$ depend on k, to lighten notation we do not make this dependence explicit. Because ES_{1-p} involves an average over the left tail containing probability p, we also define a transformed weight vector $\mathbf{w}'$:

$$w_i' := \begin{cases} -w_i/p, & i = 1, 2, \ldots, \ell \\ 0, & \text{otherwise.} \end{cases}$$

If the vector $\mathbf{V} := (V_1, V_2, \ldots, V_k)$ of true portfolio values were known, then with a weight vector $\mathbf{w}$ it would define a discrete distribution $F_{\mathbf{w},\mathbf{V}}$ assigning probability mass w_i to each value V_i. For this distribution, ES_{1-p} is $\sum_{i=1}^{k} w'_i V_{\pi_V(i)}$. The empirical likelihood confidence interval for ES_{1-p} of the unknown true distribution F_V, expressed in Equation (1), is

$$\left[\min_{\mathbf{w}\in\mathcal{S}(k)} \sum_{i=1}^{k} w'_i V_{\pi_V(i)},\ \max_{\mathbf{w}\in\mathcal{S}(k)} \sum_{i=1}^{k} w'_i V_{\pi_V(i)}\right], \tag{7}$$

representing the outer-level uncertainty entailed by working with a sample $Z_1, Z_2, \ldots, Z_k$ instead of the true distribution F_Z.

Because we do not know the values $\mathbf{V}$, we must combine this confidence interval with inner-level simulation as discussed in Lan et al. (2007b). The result, derived in Appendix A, is that the lower confidence limit is

$$\min_{\ell=\lfloor kp\rfloor,\ldots,\ell_{\max}} \left(\min_{\mathbf{w}\in\mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w'_i \bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - z_{\mathrm{lo}}(\ell) B_0(\ell)\right) \tag{8}$$

and the upper confidence limit is

$$\max_{\ell=\ell_{\min},\ldots,\lceil kp\rceil} \left(\max_{\mathbf{w}\in\mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w'_i \bar{X}_{\pi_1(i)}(N_{\pi_1(i)}) + z_{\mathrm{hi}} B_S(\ell)\right), \tag{9}$$

where several quantities are defined in Step 5 of the procedure in the following subsection. At an intuitive level, the lower confidence limit in Equation (8) arises from that in Equation (7) by ordering the scenarios based on information available at the end of the first stage, estimating the scenarios' values by second-stage sample averages, and subtracting a term that accounts for inner-level uncertainty. The upper confidence limit in Equation (9) arises similarly, but the ordering of the scenarios is based on information available at the end of the second stage, and we add a different term to account for inner-level uncertainty. The minimization and maximization over ℓ represent our uncertainty about how many of the values $V_1, V_2, \ldots, V_k$ are less than the quantile v_p.

3.3 The Procedure

The procedure involves a fixed computational budget, which may be expressed as a total number C of simulation replications, i.e., the total number of payoffs that can be simulated, or as an amount of computing time T. The distinction between these two kinds of computational budgets is important when choosing the first-stage sample size n_0 and the number of scenarios k. However, given n_0 and k, the kind of budget makes only a small difference in determining the number C_1 of payoffs to simulate in the second stage. If the budget is C total payoffs, then $C_1 = C - kn_0$. If the budget is an amount of time T, then during the first stage we must estimate t, the amount of time required to simulate one payoff, and record T_0, the amount of time required by the first stage. Then $C_1 = (T - T_0)/t$, treating the amount of time required to construct the confidence interval at the end as negligible in comparison to simulating payoffs.

To explain exactly how CRN are used, we overload notation by supposing that there is a function $X_i(\cdot)$ such that when U is a uniform random variate (or vector), the distribution of $X_i(U)$ is the conditional distribution of the payoff X given that the scenario is Z_i.

The procedure has the following steps:

1. **Scenario Generation**:
   Generate scenarios $Z_1, Z_2, \ldots, Z_k$ independently from the distribution F_Z.

2. **First Stage Sampling**:
   Sample $U_1, U_2, \ldots, U_{n_0}$ independently from a uniform distribution.
   For each $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_0$, compute $X_{ij} := X_i(U_j)$.

3. **Screening**:
   For each $i = 1, 2, \ldots, k$, compute the sample average $\bar{X}_i(n_0)$ and sample variance $S_i^2(n_0)$ of $X_{i1}, X_{i2}, \ldots, X_{in_0}$.
   Sort to produce a permutation π_0 of $\{1, 2, \ldots, k\}$ such that $\bar{X}_{\pi_0(i)}(n_0)$ is nondecreasing in i.
   Compute $\tilde{S}^2(n_0) := \max_{i=1,2,\ldots,\lceil kp \rceil} S^2_{\pi_0(i)}$ and d according to Equation (5).
   Initialize $I \leftarrow \{1, 2, \ldots, \lceil kp \rceil\}$ and $i \leftarrow k$.

   (a) *Pre-screening*: If $\bar{X}_{\pi_0(i)} > \bar{X}_{\pi_0(\lceil kp \rceil)} + d\sqrt{(S^2_{\pi_0(i)}(n_0) + \tilde{S}^2(n_0))/n_0}$, scenario $\pi_0(i)$ is pre-screened out: go to Step (3c).

   (b) *Screening*: Initialize $b \leftarrow 0$ and $j \leftarrow 1$.

      i. Compute the sample variance $S^2_{\pi_0(i)\pi_0(j)}$ of $X_{\pi_0(i)1} - X_{\pi_0(j)1}, X_{\pi_0(i)2} - X_{\pi_0(j)2}, \ldots, X_{\pi_0(i)n_0} - X_{\pi_0(j)n_0}$.

      ii. If $\bar{X}_{\pi_0(i)}(n_0) > \bar{X}_{\pi_0(j)}(n_0) + dS_{\pi_0(i)\pi_0(j)}/\sqrt{n_0}$, scenario $\pi_0(i)$ beats scenario $\pi_0(j)$: set $b \leftarrow b + 1$.

      iii. If $b \geq \lceil kp \rceil$, scenario $\pi_0(i)$ is screened out: go to Step (3c).
      Otherwise, set $j \leftarrow j + 1$.

      iv. If $j < i$, go to Step 3(b)i.
      Otherwise, scenario $\pi_0(i)$ survives screening: set $I \leftarrow I \cup \{\pi_0(i)\}$.

   (c) *Loop*: Set $i \leftarrow i - 1$. If $i > \lceil kp \rceil$, go to Step (3a).

4. **Restarting and Second Stage Sampling**:
   Discard all payoffs from Step 2.
   Compute $C_1 := C - kn_0$ or $(T - T_0)/t$, depending on the type of budget constraint.
   For each $i \in I$, compute
   $$N_i := \left\lceil \frac{C_1 S_i^2(n_0)}{\sum_{j \in I} S_j^2(n_0)} \right\rceil. \tag{10}$$
   For each $i \in I$ and $j = 1, 2, \ldots, N_i$, sample U_{ij} independently from a uniform distribution and compute $X_{ij} := X_i(U_{ij})$.

5. **Constructing the Confidence Interval**:
For each $i \in I$, compute the sample average $\bar{X}_i(N_i)$ and sample variance $S_i^2(N_i)$ of $X_{i1}, X_{i2}, \ldots, X_{iN_i}$, and compute $s_i := \sqrt{S_i^2(N_i)/N_i}$.
Compute

$$\ell_{\min} := \min\left\{\ell : k^k \left(\frac{p}{\ell}\right)^{\ell} \left(\frac{1-p}{k-\ell}\right)^{k-\ell} \geq c\right\} \quad \text{and}$$
$$\ell_{\max} := \max\left\{\ell : k^k \left(\frac{p}{\ell}\right)^{\ell} \left(\frac{1-p}{k-\ell}\right)^{k-\ell} \geq c\right\}.$$

Initialize $\hat{L} \leftarrow \infty$, $N_{\text{lo}}(\lceil kp \rceil) := \min_{i=1,2,\ldots,\lceil kp \rceil} N_{\pi_0(i)}$, and $\underline{s}(\lceil kp \rceil) := \max_{i=1,2,\ldots,\lceil kp \rceil} s_{\pi_0(i)}$.
For $\ell = \lfloor kp \rfloor, \lfloor kp \rfloor + 1, \ldots, \ell_{\max}$,

(a) Compute $z_{\text{lo}}(\ell) := t_{1-\alpha_{\text{lo}}, N_{\text{lo}}(\ell)-1}$,

$$\Delta(\ell) := \sqrt{\max_{w \in \mathcal{S}(\ell)} \sum_{i=1}^{\ell} (w_i')^2}, \tag{11}$$

and $B_0(\ell) := \underline{s}(\ell)\Delta(\ell)$.

(b) Set

$$\hat{L} \leftarrow \min\left\{\hat{L}, \min_{\mathbf{w} \in \mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' \bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - z_{\text{lo}}(\ell) B_0(\ell)\right\}. \tag{12}$$

(c) Compute $N_{\text{lo}}(\ell+1) := \min\{N_{\text{lo}}(\ell), N_{\pi_0(\ell+1)}\}$ and $\underline{s}(\ell+1) \leftarrow \max\{\underline{s}(\ell), s_{\pi_0(\ell+1)}\}$.

Sort to produce a mapping π_S from $\{1, 2, \ldots, |I|\}$ to I such that $s_{\pi_s(i)}$ is nonincreasing in i.
Initialize $\hat{U} \leftarrow -\infty$ and $\bar{s}(\ell_{\min}) := \max_{i=1,2,\ldots,\ell_{\min}} s_{\pi_s(i)}$.
Sort to produce a mapping π_1 from $\{1, 2, \ldots, |I|\}$ to I such that $\bar{X}_{\pi_1(i)}(N_{\pi_1(i)})$ is non-decreasing in i.
Compute $N_{\text{hi}} := \min\{N_{\pi_1(1)}, N_{\pi_1(2)}, \ldots, N_{\pi_1(|I|)}\}$ and $z_{\text{hi}} := t_{1-\alpha_{\text{hi}}, N_{\text{hi}}-1}$.
For $\ell = \ell_{\min}, \ell_{\min}+1, \ldots, \lceil kp \rceil$,

(a) Compute $\Delta(\ell)$ as in Equation (11) and $B_S(\ell) := \bar{s}(\ell)\Delta(\ell)$.

(b) Set

$$\hat{U} \leftarrow \max\left\{\hat{U}, \max_{\mathbf{w} \in \mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' \bar{X}_{\pi_1(i)}(N_{\pi_1(i)}) + z_{\text{hi}} B_S(\ell)\right\}. \tag{13}$$

(c) Set $\bar{s} \leftarrow \max\{\bar{s}, s_{\pi_s(\ell+1)}\}$.

The confidence interval given in Equations (8) and (9) is $[\hat{L}, \hat{U}]$.

The maximum in Equation (11) is computed by using the Newton method to solve its KKT conditions. An algorithm for the optimizations in Equations (12) and (13) is given in Baysal and Staum (2008). The t-quantiles $z_{\text{lo}}(\ell)$ and z_{hi} may be replaced by normal quantiles when the second-stage sample sizes are sufficiently large, as they typically are.

4 Experimental Results

We tested the simulation procedures by producing a 90% confidence interval (CI) for $\mathrm{ES}_{0.99}$ in the examples described in Section 2. The error $\alpha = 10\%$ was decomposed into $\alpha_o = 5\%$ for the outer level, $\alpha_s = 2\%$ for screening, and $\alpha_{\mathrm{lo}} = \alpha_{\mathrm{hi}} = 1.5\%$ for the inner-level lower and upper confidence limits. In each experiment, we chose our procedure's parameters k and n_0 according to a method described in Lan (2009). We compare our procedure with the `plain` procedure, a one-stage procedure that does not use screening, and therefore does not use CRN. It assigns an equal number of replications to each scenario, C/k if the total number C of replications is fixed, T/tk if the total computation time T is fixed. It then computes the confidence interval in Equations (8) and (9). We ran the plain procedure with the same number k of scenarios as our procedure. To compare the procedures, we evaluate their confidence intervals' coverage rates and mean widths given the same fixed budget. We ran the experiments on a PC with a 2.4 GHz CPU and 4 GB memory under 64-bit Red Hat Linux. The code was written in C++ and compiled by `gcc` 3.4.6.

Similar to results reported in Lan et al. (2007a), we found that the plain procedure and our procedure both had coverage rates greater than the nominal confidence level of 90% as long as $k \geq 40/p$, where p is the tail probability under consideration. In these examples, $p = 1 - 0.99 = 0.01$.

The following figures report average CI widths for 20 independent runs of the procedures. The error bars in the figures provide 95% confidence intervals for the mean width of our procedure's CI. (The width of the CI produced by the plain procedure is less variable, so the error bars for the plain procedure were too small to display.) In each figure, a horizontal line represents 10% relative error, that is, its value is one tenth of $\mathrm{ES}_{0.99}$. We include the line for the purpose of comparing the CI widths to a rough measure of desirable precision. It would not be very useful to attain a relative error far less than 10% because of model risk: that is, risk management models are not generally accurate enough that precision better than, say, 1% would convey meaningful information. On the other hand, if the CI width is much greater than 10% relative error, then the simulation experiment has left us with a great deal of uncertainty about the magnitude of $\mathrm{ES}_{0.99}$. For these reasons, we ran experiments with computational budgets such that our procedure yields CI widths in the neighborhood of 10% relative error.

Figure 1 shows how average CI width varies with a computational budget of C replications for the example of selling a put option described in Section 2.1. The much narrower CI widths achieved by our procedure show that the benefit of screening in directing more replications to important scenarios outweighs the cost of restarting and throwing out first-stage replications. In these experiments, our procedure produced a CI up to 116 times narrower than that produced by the plain procedure. On the log-log plot in Figure 1, the CI width decreases roughly linearly in the budget, with slope about -0.4 or -0.44. This is unfavorable compared to the usual $\mathcal{O}(C^{-1/2})$ order of convergence of ordinary Monte Carlo, but favorable compared to the $\mathcal{O}(C^{-1/3})$ order of convergence for a two-level simulation estimator of VaR found by Lee (1998) or the $\mathcal{O}(C^{-1/4})$ order of convergence for the procedure we proposed in Lan et al.

(2007a).

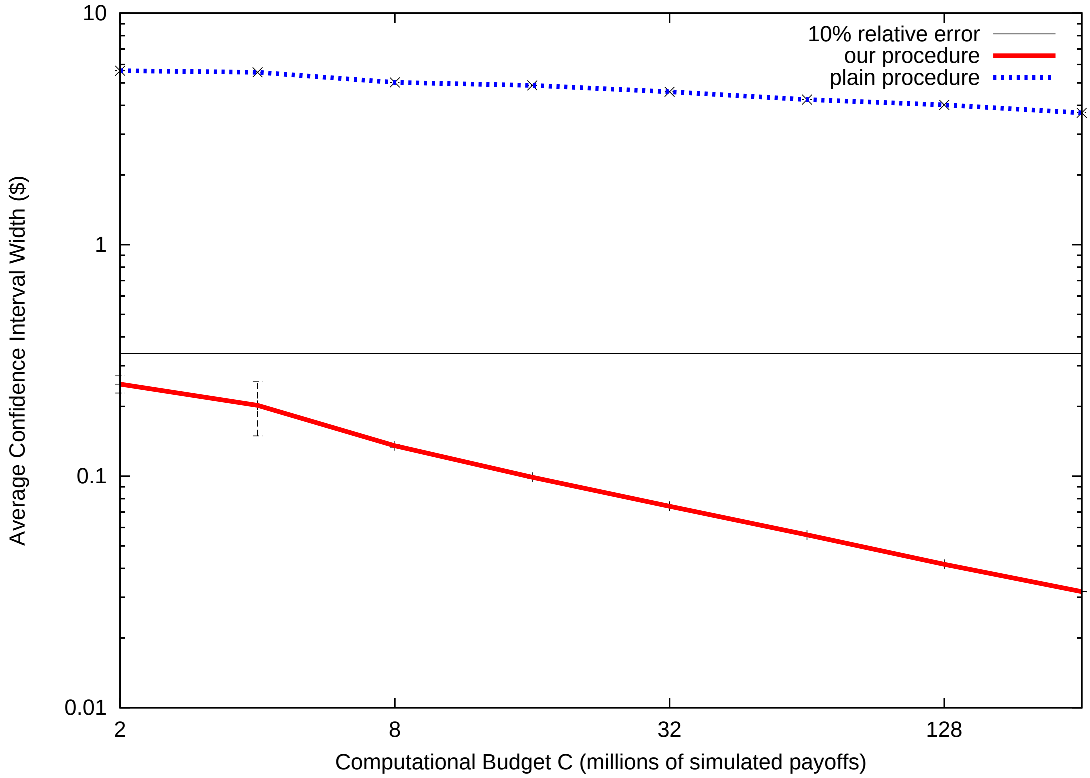

Figure 1: Average confidence interval width in the example of Section 2.1 given a fixed budget of simulation replications.

Figure 2 shows similar results from the example of an options portfolio described in Section 2.2. In this example, larger computational budgets are required to get an precise estimate of ES. Again, our procedure produced CIs narrower than those from the plain procedure, up to a factor of 14. For low budgets, our procedure's advantage was not as great. For example, when C is 32 million, our best choice was $k = 4000$ and $n_0 = 4703$, so that more than half the budget was used up in the first stage before restarting, yet the first stage was too small to enable the procedure to screen out most of the scenarios that do not belong to the tail. When the computational budget is small, our procedure may not be able to produce a CI narrow enough to be useful. A multi-stage screening procedure (similar to Lesnevski et al., 2007, 2008) might overcome this problem.

Next we present results when the computational budget limits computing time. The budget constraint is implemented not by dynamically terminating the procedures when a given amount of clock time has elapsed, but by choosing values of k and n_0 such that the procedure takes approximately the given amount of time. Our procedure's running time is slightly variable, but all experiments' durations were within 5% of the allotted time. A

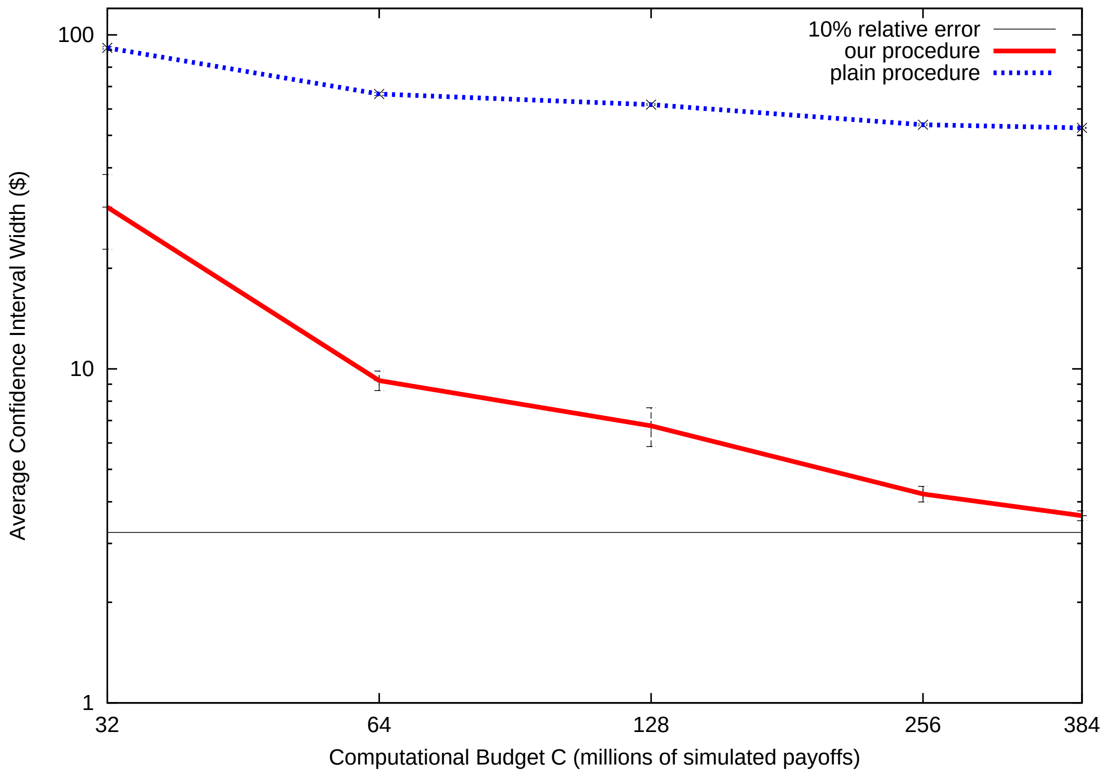

Figure 2: Average confidence interval width in the example of Section 2.2 given a fixed budget of simulation replications.

budget expressed in computing time is less favorable to our procedure (relative to the plain procedure) than a budget for the total number of replications: our procedure can spend a substantial amount of time in performing comparisons between scenarios as part of screening, even though it does not generate more replications then. The amount of time spent on screening when there are k scenarios is $\mathcal{O}(k^2)$ because there are $k^2/2$ pairs of scenarios that can be compared. This pushes us to choose smaller values of k (Lan, 2009). For example, in the example of a single put option (Section 2.1), our procedure attains a CI width around \$0.0427 with a budget of $C = 120$ million replications or $T = 1,560$ seconds, but if the budget is in replications then we choose k to be about 600,000 scenarios, whereas if the budget is in computing time, we choose k to be about 427,000 scenarios. For budgets so large as to lead to choosing a very large k, the advantage of our procedure degrades. This can be seen in Figure 3, where the curve representing our procedure's CI width becomes flatter as the computing time T grows. Still, Figures 3 and 4 show that our procedure performs much better than the plain procedure when they are given equal computing times, producing a CI narrower by a factor of as much as 15 or 12 in these two examples.

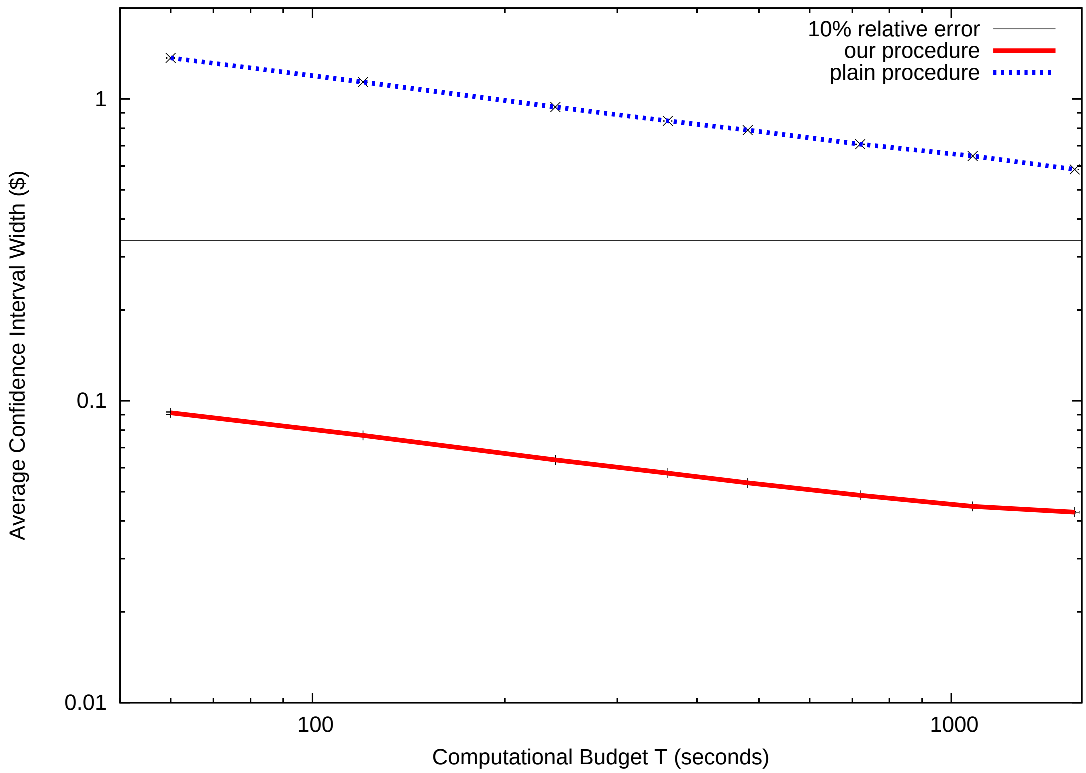

Figure 3: Average confidence interval width in the example of Section 2.1 given a fixed budget of computing time.

5 Conclusions and Future Research

We have presented and tested a new two-level simulation procedure that creates an asymptotically valid confidence interval for expected shortfall given a computational budget expressed in computing time or total number of simulation replications. We found that the confidence interval has adequate coverage as long as the number of simulated scenarios $k \geq 40/p$, where p is the tail probability at which expected shortfall is measured. In these examples, our procedure's confidence interval was dozens of times narrower than one created without using our efficiency techniques.

There are several possibilities for further improving the procedure's efficiency. Baysal and Staum (2008) mentions potential enhancements to the empirical likelihood estimation used here. As in Lesnevski et al. (2008), it may help to use multi-stage screening and to employ other inner-level variance reduction techniques, such as control variates. One might also apply variance reduction techniques at the outer level, in sampling scenarios. Relevant ideas are described by Glasserman (2004, Ch. 9); they apply to expected shortfall as well as value-at-risk. However, it seems more difficult to employ variance reduction at the outer level than the inner level while maintaining validity of the confidence interval, which is based

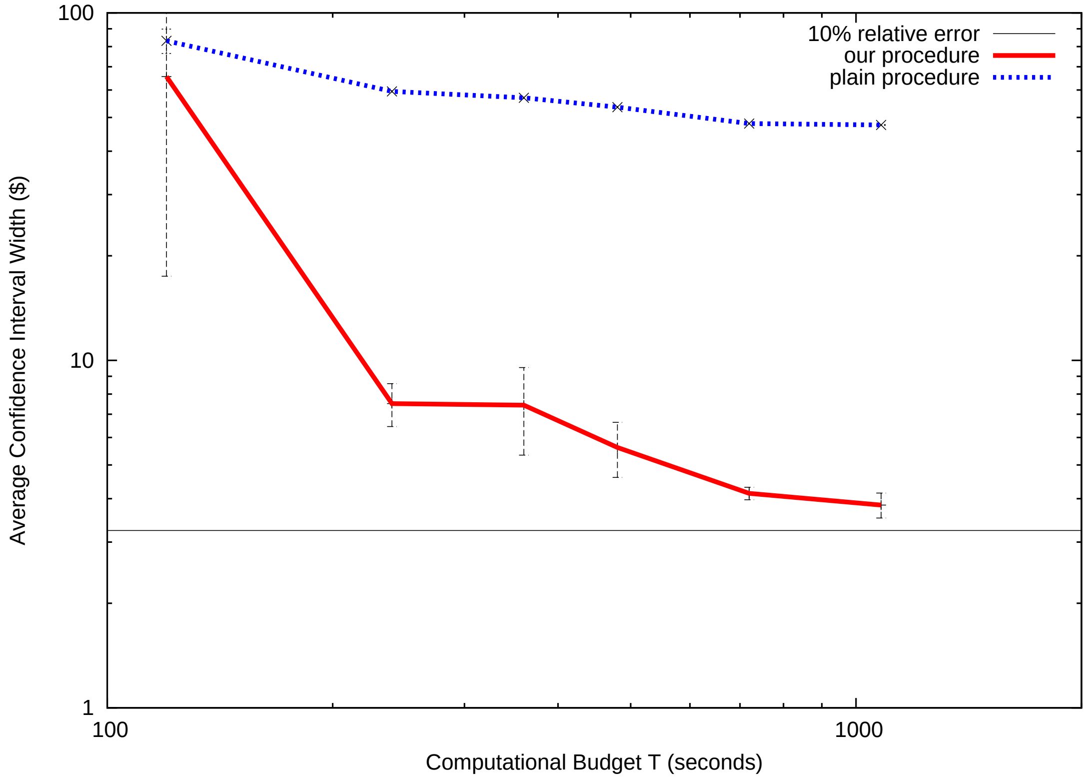

Figure 4: Average confidence interval width in the example of Section 2.2 given a fixed budget of computing time.

on empirical likelihood at the outer level. Furthermore, some of these variance reduction techniques for risk management may be substitutes rather than complements for our techniques: for example, importance sampling is often used in risk management simulations to increase the proportion of simulated scenarios that lead to large losses, but our procedure accomplishes something similar by screening out those scenarios that do not lead to large losses.

We tested our procedure on small examples, using a desktop computer. To be useful for large examples, the procedure must be run in a high-performance parallel computing framework. We are currently developing parallel implementations of the procedure.

To use our procedure requires choosing a computational budget and a confidence level $1 - \alpha$, decomposing α into several components that govern various sources of error, and choosing the number k of scenarios and the first-stage sample size n_0. In our experience, it is easy to decompose α in a way that makes the procedure efficient: the values we chose in Section 4 are broadly effective, a finding which agrees with Lesnevski et al. (2007). However, it is not so easy to choose k and n_0; the procedure's efficiency depends strongly on these choices, and the best choices are problem-dependent. A multi-stage procedure may make it easier to choose n_0 (Lesnevski et al., 2007, 2008). Ways of choosing k and n_0, either from a pilot experiment or based on experience in performing similar risk management simulations in the recent past, are the subject of ongoing research and will be discussed in Lan (2009).

References

Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. Journal of Banking and Finance 26, 1487–1503.

Banerjee, S., 1961. On confidence interval for two-means problem based on separate estimates of variances and tabulated values of t-table. Sankhyā A23, 359–378.

Baysal, R. E., Staum, J., 2008. Empirical likelihood for value at risk and expected shortfall. Journal of Risk 11 (1).

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

Chen, S. X., 2008. Nonparametric estimation of expected shortfall. Journal of Financial Econometrics 6 (1), 87–107.

Derman, E., 1999. Regimes of volatility. Risk 12 (4), 55–59.

Glasserman, P., 2004. Monte Carlo Methods in Financial Engineering. Springer-Verlag, New York.

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

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. Ph.D. thesis in preparation, Dept. of IEMS, Northwestern University.

Lan, H., Nelson, B. L., Staum, J., 2007a. A confidence interval for tail conditional expection via two-level simulation. In: Henderson, S. G., Biller, B., Hsieh, M.-H., Shortle, J., Tew, J. D., Barton, R. R. (Eds.), Proceeedings of the 2007 Winter Simulation Conference. IEEE Press, pp. 949–957.

Lan, H., Nelson, B. L., Staum, J., 2007b. 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.

Law, A. M., Kelton, W. D., 2000. Simulation Modeling and Analysis, 3rd Edition. McGraw-Hill, New York.

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

Lesnevski, V., Nelson, B. L., Staum, J., 2007. Simulation of coherent risk measures based on generalized scenarios. Management Science 53 (11), 1756–1769.

Lesnevski, V., Nelson, B. L., Staum, J., 2008. An adaptive procedure for estimating coherence risk measures based on generalized scenarios. Journal of Computational Finance 11 (4), 1–31.

Luenberger, D. G., 1998. Investment Science. Oxford University Press, New York.

McNeil, A. J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management. Princeton University Press, Princeton, New Jersey.

Owen, A. B., 2001. Empirical Likelihood. Chapman & Hall/CRC, Boca Raton.

Steckley, S. G., Henderson, S. G., 2003. Simulation input modeling: a kernel approach to estimating the density of a conditional expectation. In: Chick, S. E., Sanchez, P. J., Ferrin, D. M., Morrice, D. J. (Eds.), Proceeedings of the 2003 Winter Simulation Conference. IEEE Press, pp. 383–391.

A Proof

The procedure in Section 3.3 yields a confidence interval $[\hat{L}, \hat{U}]$ where $\hat{L}$ is defined by Equation (8) and $\hat{U}$ is defined by Equation (9). This section presents a proof that $[\hat{L}, \hat{U}]$ has asymptotic coverage at least $1 - \alpha$ as the number of scenarios $k \to \infty$, if the payoffs are normally distributed and pre-screening is omitted. In reality and in the example on which our experiments are run, payoffs are not normally distributed. However, sample averages of payoffs are approximately normal because of the central limit theorem if the sample sizes n_0 and N_i are large enough.

Theorem 1 *Suppose that for any scenario Z, the conditional distribution of the payoff X given Z is normal. Let $[\hat{L}, \hat{U}]$ represent the CI written in Equations (8) and (9), as produced by the procedure with pre-screening (Step 3a) omitted. Then* $\lim_{k\to\infty} \Pr\{\hat{L} \le ES_{1-p} \le \hat{U}\} \ge 1 - \alpha$.

The proof is within the framework of Lan et al. (2007b) for showing the asymptotic validity of a confidence interval generated by two-level simulation. Baysal and Staum (2008) show that the asymptotic probability that ES_{1-p} is contained in the outer-level confidence interval of Equation (7) is at least $1 - \alpha_o$ as $k \to \infty$. By the results of Lan et al. (2007b), it then suffices to construct a confidence region $\mathcal{V}$ such that

1. the probability that $\mathcal{V}$ contains the vector of true values $\mathbf{V} = (V_1, V_2, \ldots, V_k)$ is at least $1 - \alpha_i$, and

2. the two-level simulation confidence limits arise as follows:

$$\hat{L} = \min_{\mathbf{v} \in \mathcal{V}, \mathbf{w} \in \mathcal{S}(k)} \sum_{i=1}^{k} w_i' v_{\pi_v(i)} \quad \text{and} \quad \hat{U} = \max_{\mathbf{v} \in \mathcal{V}, \mathbf{w} \in \mathcal{S}(k)} \sum_{i=1}^{k} w_i' v_{\pi_v(i)}, \tag{14}$$

where $\mathcal{S}(k)$ is defined in Equation (6) and π_v is a permutation of $\{1, 2, \ldots, k\}$ such that $v_{\pi_v(i)}$ is nondecreasing in i.

We first describe $\mathcal{V}$. Define the t-quantiles

$$z_{\text{lo}}(\ell) := t_{1-\alpha_{\text{lo}}, \min_{i=\pi_0(1),\ldots,\pi_0(\ell)}\{N_i - 1\}} \quad \text{and} \quad z_{\text{hi}} := t_{1-\alpha_{\text{hi}}, \min_{i=\pi_1(1),\ldots,\pi_1(|I|)}\{N_i - 1\}}.$$

The degrees of freedom in these formulae are the minimum degrees of freedom available in estimating any of the relevant standard deviations. Also define

$$\begin{aligned} B_0(\ell) &:= \max_{\mathbf{w} \in \mathcal{S}_\ell(k)} \sqrt{\sum_{i=1}^{\ell} (w_i')^2 \frac{S^2_{\pi_0(i)}(N_{\pi_0(i)})}{N_{\pi_0(i)}}} \quad \text{and} \\ B_S(\ell) &:= \max_{\mathbf{w} \in \mathcal{S}_\ell(k)} \sqrt{\sum_{i=1}^{\ell} (w_i')^2 \frac{S^2_{\pi_s(i)}(N_{\pi_s(i)})}{N_{\pi_s(i)}}} \end{aligned}$$

to be used as bounds on standard deviations of weighted averages. Our confidence region $\mathcal{V}$ for $\mathbf{V}$ is the set containing all vectors $\mathbf{v}$ such that

$$\forall i \notin I,\ v_i \geq v_{\pi_v(\lceil kp \rceil)}, \tag{15}$$

$$\min_{\ell=\lfloor kp \rfloor,\ldots,\ell_{\max}} \left(\min_{\mathbf{w} \in \mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' \bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - z_{\text{lo}}(\ell) B_0(\ell) \right) \leq \min_{\mathbf{w} \in \mathcal{S}(k)} \sum_{i=1}^{k} w_i' v_{\pi_v(i)}, \tag{16}$$

and

$$\max_{\ell=\ell_{\min},\ldots,\lceil kp \rceil} \left(\max_{\mathbf{w} \in \mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' \bar{X}_{\pi_1(i)}(N_{\pi_0(i)}) + z_{\text{hi}} B_S(\ell) \right) \geq \max_{\mathbf{w} \in \mathcal{S}(k)} \sum_{i=1}^{k} w_i' v_{\pi_v(i)}. \tag{17}$$

This construction makes Equation (14) hold: compare Equation (8) with Equation (16) and Equation (9) with Equation (17).

The event that the confidence region $\mathcal{V}$ includes the true values $\mathbf{V}$ can be understood by plugging in $\mathbf{V}$ for $\mathbf{v}$ in the definition of $\mathcal{V}$. If we do so, then the constraints defining $\mathcal{V}$ take on the following interpretations.

- Equation (15) is equivalent to correct screening: $\gamma := \{\pi_V(1), \pi_V(2), \ldots, \pi_V(\lceil kp \rceil)\} \subseteq I$.

- Equation (16) implies that the two-level lower confidence limit $\hat{L} \leq \min_{\mathbf{w} \in \mathcal{S}(k)} \sum_{i=1}^{k} w_i' V_{\pi_V(i)}$, the outer-level lower confidence limit.

- Equation (17) implies that the two-level upper confidence limit $\hat{U} \geq \max_{\mathbf{w} \in \mathcal{S}(k)} \sum_{i=1}^{k} w_i' V_{\pi_V(i)}$, the outer-level upper confidence limit.

By the Bonferroni inequality, the probability that $\mathbf{V} \notin \mathcal{V}$ is bounded above by the sum of the probabilities

- that $\mathbf{V}$ does not satisfy Equation (15),
- that $\mathbf{V}$ does not satisfy Equation (16), and
- that $\mathbf{V}$ satisfies Equation (15) and does not satisfy Equation (17).

In Appendix A.1 we will show that the probability that $\mathbf{V}$ does not satisfy Equation (15) is bounded above by α_s. In Appendix A.2 we will show that the other two probabilities are bounded above by α_{lo} and α_{hi} respectively. Because $\alpha_i = \alpha_s + \alpha_{\mathrm{lo}} + \alpha_{\mathrm{hi}}$, this proves that $\Pr\{\mathbf{V} \in \mathcal{V}\} \geq 1 - \alpha_i$.

A.1 Screening

Here we show that the probability of correct screening $\Pr\{\gamma \subseteq I\} \geq 1 - \alpha_s$, where $\gamma = \{\pi_V(1), \cdots, \pi_V(\lceil kp \rceil)\}$. Let

$$B_{ij} := \mathbf{1}\{\bar{X}_i(n_0) > \bar{X}_j(n_0) + dS_{ij}(n_0)/\sqrt{n_0}\}$$

be the indicator function which is 1 when scenario j beats scenario i. We have

$$\Pr\{\gamma \subseteq I\} \geq \Pr\{\forall i \in \gamma, j \notin \gamma,\ B_{ij} = 0\} \geq 1 - \sum_{i \in \gamma} \sum_{j \notin \gamma} \Pr\{B_{ij} = 1\}$$

by the Bonferroni inequality. For $i \in \gamma$ and $j \notin \gamma$, $V_i \leq V_j$. Therefore each

$$\Pr\{B_{ij} = 1\} = \Pr\left\{\frac{\bar{X}_i(n_0) - \bar{X}_j(n_0)}{S_{ij}(n_0)/\sqrt{n_0}} > d\right\} \leq \frac{\alpha_s}{\lceil kp \rceil (k - \lceil kp \rceil)},$$

using $d = t_{n_0-1, 1-\alpha_s/(k-\lceil kp \rceil)\lceil kp \rceil}$.

A.2 Confidence Region

In this section, we deal with the second-stage inner-level simulation, after screening and restarting have occurred. We can think of the first stage as randomly generating a simulation problem which the second stage solves. The first stage produces I and N_i for each $i \in I$. This is an experimental design for the second stage, specifying which scenarios to consider and how many payoffs to simulate from each of them.

To prove $\mathcal{V}$ is a confidence region for $\mathbf{V}$, we first prove two lemmas. The first lemma addresses the following issue. On the left side of Equation (16), the minimization is over $\ell \in \{\lfloor kp \rfloor, \ldots, \ell_{\max}\}$ and then over $\mathbf{w} \in \mathcal{S}_\ell(k)$. On the right side, the minimization is over $\mathbf{w} \in \mathcal{S}(k)$, which is equivalent to minimization over $\ell \in \{\ell_{\min}, \ldots, \ell_{\max}\}$ and then over $\mathbf{w} \in \mathcal{S}_\ell(k)$, according to the definition of $\mathcal{S}(k)$ in Equation (6). We formulated Equation (16) in this way so that the procedure can save time by minimizing over the smaller range $\ell \in \{\lfloor kp \rfloor, \ldots, \ell_{\max}\}$ instead of $\{\ell_{\min}, \ldots, \ell_{\max}\}$. For the same reason, in Equation (17) on the left side, we maximized over $\ell \in \{\ell_{\min}, \ldots, \lceil kp \rceil\}$ instead of over $\{\ell_{\min}, \ldots, \ell_{\max}\}$. This formulation of the confidence region complicates the proof of Theorem 2, in which we use Lemma 1 to show that we can minimize and maximize over these smaller ranges and still get the desired coverage for the confidence region.

Lemma 1 *For any k-vector v and $\ell \in \{1, 2, \ldots, \lfloor kp \rfloor\}$, there exists $\ell' \geq \lfloor kp \rfloor$ such that*

$$\min_{\mathbf{w} \in \mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' v_{\pi_v(i)} \geq \min_{\mathbf{w} \in \mathcal{S}_{\ell'}(k)} \sum_{i=1}^{\ell'} w_i' v_{\pi_v(i)}.$$

Similarly, for any k-vector v and $\ell \in \{\lceil kp \rceil, \lceil kp \rceil + 1, \ldots, k\}$, there exists $\ell' \leq \lceil kp \rceil$ such that

$$\max_{\mathbf{w} \in \mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' v_{\pi_v(i)} \leq \max_{\mathbf{w} \in \mathcal{S}_{\ell'}(k)} \sum_{i=1}^{\ell'} w_i' v_{\pi_v(i)}.$$

Proof: We prove the first part of the lemma in detail. For any $\ell \in \{1, 2, \ldots, \lfloor kp \rfloor\}$, define

$$\underline{w}(\ell) := \left(w_{\ell,1}, w_{\ell,2}, \ldots, w_{\ell,\ell}, \frac{1-p}{k-\ell}, \ldots, \frac{1-p}{k-\ell} \right)$$

where $w_{\ell,1}, w_{\ell,2}, \ldots, w_{\ell,\ell}$ are chosen so that

$$\sum_{i=1}^{\ell} w_{\ell,i}' v_{\pi_v(i)} = \min_{w \in \mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' v_{\pi_v(i)}.$$

By optimality, $w_{\ell,1} \leq w_{\ell,2} \leq \ldots \leq w_{\ell,\ell}$. Because $\ell < kp$ and the weights sum to 1, one of the weights must exceed $(1-p)/(k-\ell)$, which is less than $1/k$. Therefore there is an integer s between 0 and ℓ such that $w_{\ell,i} < (1-p)/(k-\ell)$ for all $i = 1, \ldots, s$ and $w_{\ell,i} \geq (1-p)/(k-\ell)$ for all $i = s+1, \ldots, \ell$.

First we consider the case in which the tail probability $p < 0.5$. The following construction shows there is an integer $\ell' \geq \lfloor kp \rfloor$ and a weight vector $w(\ell') \in \mathcal{S}_{\ell'}(k)$, whose ith component is to be denoted $w_{\ell',i}$, such that $\sum_{i=1}^{\ell} w_i' v_{\pi_v(i)} \geq \sum_{i=1}^{\ell'} w_{\ell',i}' v_{\pi_v(i)}$.

1. Choose ℓ' to be the largest integer such that $\sum_{i=1}^{s} w_{\ell,i} + (\ell' - s)(1-p)/(k-\ell) \leq p$. Because $\ell < \lfloor kp \rfloor$ (so $(1-p)/(k-\ell) < 1/k$) and $p < 0.5$, and from the definition of s, it follows that $(k-\ell) > \ell' \geq \lfloor kp \rfloor > \ell$. Initialize $w(\ell')$ to

$$\left(w_{\ell,1}, \ldots, w_{\ell,s}, \frac{1-p}{k-\ell}, \ldots, \frac{1-p}{k-\ell}, w_{\ell,s+1}, \ldots, w_{\ell,\ell} \right).$$

   That is, this weight vector is derived from $\underline{w}$ by switching the weights of components $s+1, \ldots, \ell'$ with the weights of components $k - \ell' + s + 1, \ldots, k$. This weight vector may not be in $\mathcal{S}_{\ell'}(k)$ because $\sum_{i=1}^{s} w_{\ell,i} + (\ell' - s)(1-p)/(k-\ell)$ may be less than p. The following steps adjust $w(\ell')$ to make $\sum_{i=1}^{\ell'} w_{\ell',i} = p$. Initialize $m \leftarrow 0$.

2. Do while $\sum_{i=1}^{\ell'} w_{\ell',i} + (w_{\ell,\ell-m} - (1-p)/(k-\ell)) \leq p$:

   - Switch the weights in components $\ell' - m$ and $k - m$: let $w_{\ell',\ell'-m} \leftarrow w_{\ell,\ell-m}$ and $w_{\ell',k-m} \leftarrow (1-p)/(k-\ell)$.
   - Set $m \leftarrow m + 1$, end while.

3. Because of the termination criterion, $\sum_{i=1}^{\ell} w_{\ell,i} = p$, and $\ell' > \ell$, we terminate with $m < \ell - s$ and $\ell' - s - m \geq 2$. Then there exists $\Delta \geq 0$ satisfying $w_{\ell,\ell-m} - \Delta > (1-p)/(k-\ell)$ and $\sum_{i=1}^{\ell'} w_{\ell',i} + \Delta = p$. Add Δ to component $\ell' - m$ and subtract it from component $k - m$: set $w_{\ell',\ell'-m} \leftarrow (1-p)/(k-\ell) + \Delta$ and $w_{\ell',k-m} \leftarrow w_{\ell,\ell-m} - \Delta$.

In the end, this produces a weight vector $w(\ell') \in \mathcal{S}_{\ell'}(k)$ such that $\sum_{i=1}^{\ell} w'_i v_{\pi_v(i)} \geq \sum_{i=1}^{\ell'} w'_{\ell',i} v_{\pi_v(i)}$.

In the case where $p > 0.5$, the proof is similar. In this case there is an integer s' such that $\sum_{i=1}^{s'} w_{\ell,i} + (1-p) \leq p$. Likewise, the proof of the second part of the lemma is similar, except that we will be choosing $\ell' < \ell$. $\Box$

The next lemma provides a tool like a t-test for weighted sums of independent normal random variables.

Lemma 2 *Suppose $X_{ij} \sim N(V_i, \sigma_i)$ are independent for $i = 1, \ldots, \ell$ and $j = 1, \ldots, N_i$. Let $\bar{X}_i$ and S_i^2 be respectively the sample mean and sample variance of $X_{i1}, \ldots, X_{iN_i}$. Suppose $\mathbf{w}$ is a nonnegative ℓ-vector whose elements sum to 1 and π is a permutation of $i = 1, 2, \ldots, \ell$. Define $A := \sum_{i=1}^{\ell} w_i(\bar{X}_{\pi(i)} - V_{\pi(i)})$ and $S^2 := \sum_{i=1}^{\ell} w_i^2 \left(S_{\pi(i)}^2/N_{\pi(i)}\right)$. If $0 < \epsilon < 0.5$ then $\Pr\left\{A \geq -t_{1-\epsilon,\min_{i=1,\ldots,\ell}\{N_i-1\}} S\right\} \geq 1 - \epsilon$.*

Proof: Because of the independence, A is normal with mean 0 and variance $\sum_{i=1}^{\ell} w_i^2 \sigma_{\pi(i)}^2/N_{\pi(i)}$, and it is independent of $S_1^2/N_1, S_2^2/N_2, \ldots, S_\ell^2/N_\ell$, which are themselves mutually independent. We can write

$$S^2 = \sum_{i=1}^{\ell} w_i^2 \left(\frac{S_{\pi(i)}^2}{N_{\pi(i)}}\right) = \sigma_A^2 \sum_{i=1}^{\ell} \frac{w_i^2}{\sigma_A^2} \left(\frac{\sigma_{\pi(i)}^2}{N_{\pi(i)}}\right) \left(\frac{S_{\pi(i)}^2}{\sigma_{\pi(i)}^2}\right)$$

where $\sigma_A^2 := \sum_{i=1}^{l} w_i^2 \sigma_{\pi(i)}^2/N_{\pi(i)}$. Also define $\lambda_i := (w_i^2/\sigma_A^2)(\sigma_{\pi(i)}^2/N_{\pi(i)})$. Notice that $\lambda_1, \ldots, \lambda_\ell$ are nonnegative weights that sum to 1. The distribution of $(N_{\pi(i)} - 1) S_{\pi(i)}^2/\sigma_{\pi(i)}^2$ is chi-squared with $N_{\pi(i)} - 1$ degrees of freedom. By a property of the t distribution,

$$t^2_{1-\epsilon,\min_{i=1,\ldots,\ell}\{N_i-1\}} \geq t^2_{1-\epsilon,N_i-1}$$

for all $i = 1, 2, \ldots, \ell$. Because $\epsilon < 0.5$,

$$\begin{aligned}
&\Pr\left\{A \geq -t_{1-\epsilon,\min_{i=1,\ldots,\ell}\{N_i-1\}} S\right\} \\
&= \frac{1}{2} + \frac{1}{2}\Pr\left\{\frac{A^2}{\sigma_A^2} \geq t^2_{1-\epsilon,\min_{i=1,\ldots,\ell}\{N_i-1\}} \sum_{i=1}^{\ell} \frac{w_i^2}{\sigma_A^2}\left(\frac{\sigma_{\pi(i)}^2}{N_{\pi(i)}}\right)\left(\frac{S_{\pi(i)}^2}{\sigma_{\pi(i)}^2}\right)\right\} \\
&\geq \frac{1}{2} + \frac{1}{2}\Pr\left\{\frac{A^2}{\sigma_A^2} \geq \sum_{i=1}^{\ell} t^2_{1-\epsilon,N_{\pi(i)}-1} \lambda_i \left(\frac{S_{\pi(i)}^2}{\sigma_{\pi(i)}^2}\right)\right\}.
\end{aligned}$$

By Banerjee's Theorem (Banerjee, 1961),

$$\Pr\left\{A \geq -t_{1-\epsilon,\min_{i=1,\ldots,\ell}\{N_i-1\}} S\right\} \geq \frac{1}{2} + \frac{1}{2}(1-2\epsilon) = 1-\epsilon.$$

□

Using this lemma, we can prove the main result, that $\mathcal{V}$ defined by Equations (15), (16) and (17) is a confidence region for $\mathbf{V}$ with confidence level $1-\alpha_i$, where $\alpha_i = \alpha_s + \alpha_{\text{hi}} + \alpha_{\text{lo}}$.

Theorem 2 *Suppose that for any scenario Z, the conditional distribution of the payoff X given Z is normal. If the payoffs are simulated independently, then* $\Pr\{\mathbf{V} \in \mathcal{V}\} \geq 1-\alpha_i$.

Proof: To condense notation, define $\mathcal{V}(\text{CS})$ as the set containing $\mathbf{v}$ satisfying Equation (15), $\mathcal{L}$ as the set containing $\mathbf{v}$ satisfying Equation (16), and $\mathcal{U}$ as the set containing $\mathbf{v}$ satisfying Equation (17), so that $\mathcal{V} = \mathcal{V}(\text{CS}) \cap \mathcal{L} \cap \mathcal{U}$. Then

$$\Pr\{\mathbf{V} \in \mathcal{V}\} = \Pr\{\mathbf{V} \in \{\mathcal{V}(\text{CS}) \cap \mathcal{L} \cap \mathcal{U}\}\} \geq 1 - \Pr\{\mathbf{V} \notin \mathcal{L}\} - \Pr\{\mathbf{V} \notin \{\mathcal{U} \cap \mathcal{V}(\text{CS})\}\}.$$

Also define

$$\underline{\ell} := \arg\min_{\ell=\lfloor kp\rfloor,\ldots,\ell_{\max}} \left(\min_{w\in\mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' V_{\pi_V(i)}\right) \quad \text{and} \quad \underline{w} := \arg\min_{w\in\mathcal{S}_{\underline{\ell}}(k)} \sum_{i=1}^{\underline{\ell}} w_i' V_{\pi_0(i)}.$$

By Lemma 1,

$$\min_{\mathbf{w}\in\mathcal{S}(k)} \sum_{i=1}^{k} w_i' V_{\pi_V(i)} = \min_{w\in\mathcal{S}_{\underline{\ell}}(k)} \sum_{i=1}^{\underline{\ell}} w_i' V_{\pi_V(i)}. \tag{18}$$

Because $V_{\pi_V(1)}, V_{\pi_V(2)}, \ldots, V_{\pi_V(\underline{\ell})}$ are the lowest $\underline{\ell}$ values among $V_1, V_2, \ldots, V_k$, while the elements of $\mathbf{w}'$ are negative,

$$\min_{w\in\mathcal{S}_{\underline{\ell}}(k)} \sum_{i=1}^{\underline{\ell}} w_i' V_{\pi_V(i)} \geq \min_{w\in\mathcal{S}_{\underline{\ell}}(k)} \sum_{i=1}^{\underline{\ell}} w_i' V_{\pi_0(i)}. \tag{19}$$

Let $\mathcal{F}_0$ represent the information produced by the first stage of the procedure. Using Equations (18)–(19),

$$\begin{aligned}
&\Pr\{\mathbf{V} \notin \mathcal{L} | \mathcal{F}_0\} \\
&= \Pr\left\{\min_{\mathbf{w}\in\mathcal{S}(k)} \sum_{i=1}^{k} w_i' V_{\pi_V(i)} < \min_{\ell=\lfloor kp\rfloor,\ldots,\ell_{\max}} \left(\min_{\mathbf{w}\in\mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' \bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - z_{\text{lo}}(\ell)B_0(\ell)\right) \middle| \mathcal{F}_0\right\} \\
&= \Pr\left\{\min_{\mathbf{w}\in\mathcal{S}_{\underline{\ell}}(k)} \sum_{i=1}^{\underline{\ell}} w_i' V_{\pi_V(i)} < \min_{\ell=\lfloor kp\rfloor,\ldots,\ell_{\max}} \left(\min_{\mathbf{w}\in\mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' \bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - z_{\text{lo}}(\ell)B_0(\ell)\right) \middle| \mathcal{F}_0\right\} \\
&\leq \Pr\left\{\min_{\mathbf{w}\in\mathcal{S}_{\underline{\ell}}(k)} \sum_{i=1}^{\underline{\ell}} w_i' V_{\pi_0(i)} < \min_{\ell=\lfloor kp\rfloor,\ldots,\ell_{\max}} \left(\min_{\mathbf{w}\in\mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w_i' \bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - z_{\text{lo}}(\ell)B_0(\ell)\right) \middle| \mathcal{F}_0\right\} \\
&\leq \Pr\left\{\sum_{i=1}^{\underline{\ell}} \underline{w}_i' V_{\pi_0(i)} < \sum_{i=1}^{\underline{\ell}} \underline{w}_i' \bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - z_{\text{lo}}(\underline{\ell})B_0(\underline{\ell}) \middle| \mathcal{F}_0\right\} \\
&= \Pr\left\{\sum_{i=1}^{\underline{\ell}} \underline{w}_i' (\bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - V_{\pi_0(i)}) > z_{\text{lo}}(\underline{\ell})B_0(\underline{\ell}) \middle| \mathcal{F}_0\right\}.
\end{aligned}$$

Because we restart after the first stage, the conditional distribution of $\sum_{i=1}^{\ell} \underline{w}_i'(\bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - V_{\pi_0(i)})$ given $\mathcal{F}_0$ is normal with mean 0 and variance $\sum_{i=1}^{\ell}(\underline{w}_i')^2\sigma^2_{\pi_0(i)}/N_{\pi_0(i)}$. By Lemma 2,

$$\Pr\left\{\sum_{i=1}^{\ell} \underline{w}_i'(\bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - V_{\pi_0(i)}) > z_{\text{lo}}(\underline{\ell})B_0(\underline{\ell}) \middle| \mathcal{F}_0\right\} \leq \alpha_{\text{lo}}.$$

Thus $\Pr\{\mathbf{V} \notin \mathcal{L}\} = \mathrm{E}\left[\Pr\{\mathbf{V} \notin \mathcal{L} : \mathcal{F}_0\}\right] \leq \alpha_{\text{lo}}$.

Similarly,

$$\begin{aligned}
\Pr\{\mathbf{V} \notin \mathcal{U} \cap \mathcal{V}(\text{CS})\} &= \Pr\{\mathbf{V} \notin \mathcal{V}(\text{CS})\} + \Pr\{\mathbf{V} \in \mathcal{V}(\text{CS}), \mathbf{V} \notin \mathcal{U}\} \\
&\leq \alpha_s + \mathrm{E}\left[\Pr\{\mathbf{V} \notin \mathcal{U}|\mathcal{F}_0\}\mathbf{1}\{\mathbf{V} \in \mathcal{V}(\text{CS})\}\right].
\end{aligned}$$

Define

$$\bar{\ell} := \arg\max_{\ell=\ell_{\min},\ldots,\lceil kp\rceil}\left\{\max_{w\in\mathcal{S}_\ell(k)}\sum_{i=1}^{\ell} w_i' V_{\pi_V(i)}\right\} \quad \text{and} \quad \bar{w} := \arg\max_{w\in\mathcal{S}_{\bar{\ell}}(k)}\sum_{i=1}^{\bar{\ell}} w_i' V_{\pi_V(i)}.$$

By Lemma 1,

$$\max_{w\in\mathcal{S}(k)}\sum_{i=1}^{k} w_i' V_{\pi_V(i)} = \sum_{i=1}^{\bar{\ell}} \bar{w}_i' V_{\pi_V(i)}. \tag{20}$$

As shown by Baysal and Staum (2008), by optimality, $\bar{w}_1', \bar{w}_2', \ldots, \bar{w}_{\bar{\ell}}'$ is negative and non-decreasing, so

$$\sum_{i=1}^{\bar{\ell}} \bar{w}_i'\bar{X}_{\pi_1(i)}(N_{\pi_1(i)}) \geq \sum_{i=1}^{\bar{\ell}} \bar{w}_i'\bar{X}_{\pi_V(i)}(N_{\pi_V(i)}). \tag{21}$$

Using Equations (20)–(21),

$$\begin{aligned}
&\Pr\{\mathbf{V} \notin \mathcal{U}|\mathcal{F}_0\} \\
&= \Pr\left\{\max_{w\in\mathcal{S}(k)}\sum_{i=1}^{k} w_i' V_{\pi_V(i)} > \max_{\ell=\ell_{\min},\ldots,\lceil kp\rceil}\left\{\max_{\mathbf{w}\in\mathcal{S}_\ell(k)}\sum_{i=1}^{\ell} w_i'\bar{X}_{\pi_1(i)}(N_{\pi_1(i)}) + z_{\text{hi}}B_S(\ell)\right\} \middle| \mathcal{F}_0\right\} \\
&= \Pr\left\{\sum_{i=1}^{\bar{\ell}} \bar{w}_i' V_{\pi_V(i)} > \max_{\ell=\ell_{\min},\ldots,\lceil kp\rceil}\left\{\max_{\mathbf{w}\in\mathcal{S}_\ell(k)}\sum_{i=1}^{\ell} w_i'\bar{X}_{\pi_1(i)}(N_{\pi_1(i)}) + z_{\text{hi}}B_S(\ell)\right\} \middle| \mathcal{F}_0\right\} \\
&\leq \Pr\left\{\sum_{i=1}^{\bar{\ell}} \bar{w}_i' V_{\pi_V(i)} > \sum_{i=1}^{\bar{\ell}} \bar{w}_i'\bar{X}_{\pi_1(i)}(N_{\pi_1(i)}) + z_{\text{hi}}B_S(\bar{\ell}) \middle| \mathcal{F}_0\right\} \\
&\leq \Pr\left\{\sum_{i=1}^{\bar{\ell}} \bar{w}_i' V_{\pi_V(i)} > \sum_{i=1}^{\bar{\ell}} \bar{w}_i'\bar{X}_{\pi_V(i)}(N_{\pi_V(i)}) + z_{\text{hi}}B_S(\bar{\ell}) \middle| \mathcal{F}_0\right\} \\
&= \Pr\left\{\sum_{i=1}^{\bar{\ell}} \bar{w}_i'(\bar{X}_{\pi_V(i)}(N_{\pi_V(i)}) - V_{\pi_V(i)}) < -z_{\text{hi}}B_S(\bar{\ell}) \middle| \mathcal{F}_0\right\}.
\end{aligned}$$

The conditional distribution of $\sum_{i=1}^{\bar{\ell}} \bar{w}_i'(\bar{X}_{\pi_V(i)}(N_{\pi_V(i)}) - V_{\pi_V(i)})$ given $\mathcal{F}_0$ is normal with mean 0 and variance $\sum_{i=1}^{\bar{\ell}} (\bar{w}_i')^2 \sigma^2_{\pi_V(i)} / N_{\pi_V(i)}$. If $\mathbf{V} \in \mathcal{V}(\mathrm{CS})$, then

$$B_S^2(\bar{\ell}) = \max_{w \in \mathcal{S}(\bar{\ell})} \sum_{i=1}^{\bar{\ell}} \frac{(w_i')^2 S^2_{\pi_s(i)}}{N_{\pi_s(i)}} \geq \sum_{i=1}^{\bar{\ell}} \frac{(w_i')^2 S^2_{\pi_V(i)}}{N_{\pi_V(i)}},$$

and so by Lemma 2,

$$\Pr\left\{ \sum_{i=1}^{\bar{\ell}} \bar{w}_i'(\bar{X}_{\pi_V(i)}(N_{\pi_V(i)}) - V_{\pi_V(i)}) < -z_{\mathrm{hi}} B_S(\bar{\ell}) \middle| \mathcal{F}_0 \right\} \mathbf{1}\left\{ \mathbf{V} \in \mathcal{V}(\mathrm{CS}) \right\} \leq \alpha_{\mathrm{hi}}.$$

Therefore $\mathrm{E}\left[\Pr\left\{\mathbf{V} \notin \mathcal{U} | \mathcal{F}_0\right\} \mathbf{1}\left\{\mathbf{V} \in \mathcal{V}(\mathrm{CS})\right\}\right] \leq \alpha_{\mathrm{hi}}$, and so $\Pr\left\{\mathbf{V} \notin \mathcal{U} \cap \mathcal{V}(\mathrm{CS})\right\} \leq \alpha_s + \alpha_{\mathrm{hi}}$.

In total, $\Pr\left\{\mathbf{V} \in \mathcal{V}(\mathrm{CS}) \cap \mathcal{L} \cap \mathcal{U}\right\} \geq 1 - \alpha_{\mathrm{lo}} - \alpha_s - \alpha_{\mathrm{hi}} = 1 - \alpha_i$. □