OPERATIONS RESEARCH
Vol. 58, No. 5, September–October 2010, pp. 1481–1490
ISSN 0030-364X | EISSN 1526-5463 | 10 | 5805 | 1481

DOI 10.1287/opre.1090.0792


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

**Hai Lan**
Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200052, China,
lanhai@sjtu.edu.cn

**Barry L. Nelson, Jeremy Staum**
Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208,
{nelsonb@northwestern.edu, j-staum@northwestern.edu}

We develop and evaluate a two-level simulation procedure that produces a confidence interval for expected shortfall. The outer level of simulation generates financial scenarios, whereas the inner level estimates expected loss conditional on each scenario. Our procedure uses the statistical theory of empirical likelihood to construct a confidence interval. It also uses tools from the ranking-and-selection literature to make the simulation efficient.

*Subject classifications*: conditional value-at-risk; worst conditional expectation; tail conditional expectation; conditional tail expectation; expected shortfall; empirical likelihood; two-level simulation; simulation; design of experiments: two-level simulation; simulation; efficiency: screening methods; finance; portfolio: risk management.
*Area of review*: Simulation.
*History*: Received November 2008; revisions received April 2009, August 2009; accepted September 2009. Published online in *Articles in Advance* August 17, 2010.

## 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 firmwide 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 firmwide risk measurement simulation can take days to run on a cluster of 1,000 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 $\mathrm{VaR}_{1-p}$ 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 (ES):

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

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 (TCE) (Acerbi and Tasche 2002):

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

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 $V_1, V_2, \ldots, V_k$ 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 $V_1, V_2, \ldots, V_k$, 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 $F_V$. In 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 not analytically tractable. Furthermore, the function $V(\cdot)$ itself is not generally known in closed form; it is known in closed form only for some simple models and derivative securities. However, in most models, we can represent $V(Z) = E[X \mid Z]$ 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 (2006) studies 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. A plain two-level simulation procedure, such as that discussed in §4, can produce very wide confidence intervals given the available computational budget. To produce a narrower confidence interval given a fixed computational budget, our procedure uses screening with common random numbers and allocates sample sizes proportional to each scenario's sample variance. An electronic companion to this paper is available as part of the online version at http://or.journal.informs.org. In the electronic companion, 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 §3. Numerical results of simulation experiments in which we apply our procedures to the examples appear in §4, whereas §5 concludes and describes future research.

## 2. Motivating Examples

Risk management simulations may deal with nontrivial models and thousands of derivative securities with complicated payoffs. However, for purposes of illustration, we consider the following two simple 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

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

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

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

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 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 were 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 $T$:

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

The initial stock prices were $S_0^{(1)} = \$27.15$ for CSCO and $S_0^{(2)} = \$5.01$ for JAVA. Based on sample moments of 1,000 daily stock prices, the volatilities of CSCO and JAVA are, respectively, $\sigma^{(1)} = 32.85\%$ and $\sigma^{(2)} = 47.75\%$, whereas 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 $\mu$ 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, the sticky strike assumption does not lead to a consistent model of the underlying stock prices, so one cannot use a single simulation of the two stocks to price all the options; this makes it more natural to think of eight separate option-pricing simulations. 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 $V(Z) = E[X|Z]$, 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.

**Table 1.** Portfolio of call options.

| Index $i$ | Underlying stock | Position $\theta_i$ | Strike $K_i$ (\$) | Maturity $U_i$ | Option price (\$) | Implied volatility $\sigma_i$ (%) |
|---|---|---|---|---|---|---|
| 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 | 1,200 | 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 |

## 3. Explanation of the Procedure

This section presents a fixed-budget two-level simulation procedure for generating a confidence interval for $\mathrm{ES}_{1-p}$. 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 $\mathrm{ES}_{1-p}$ 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 $\pi_V$ is a permutation of $\{1, 2, \ldots, k\}$ such that $V_{\pi_V(1)} \leqslant V_{\pi_V(2)} \leqslant \cdots \leqslant V_{\pi_V(k)}$. That is, $V_{\pi_V(i)}$ is the $i$th 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 $n_0 \geqslant 2$ 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 $\bar{X}_i(n_0)$. Our proof of the procedure's validity depends on an assumption that the payoffs are normally distributed. Financial payoffs are often far from normally distributed, but the sample averages are usually close to normally distributed because of the central limit theorem. Lesnevski et al. (2008) investigated empirically the effect of nonnormal payoffs on a related confidence interval procedure and found that it posed no problem as long as the first-stage sample size $n_0$ was at least 30.

After the first stage, screening eliminates scenarios whose values are not likely to appear in Equation (2). The goal of screening is to allocate more of the computational budget to scenarios that matter. After screening, 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. For scenarios that survive screening, we set the sample size proportional to sample variance, with the goal of equalizing the standard errors of simulating each scenario's value. The reason for this is that the confidence interval width is related to the maximum standard error over all scenarios because of the proof techniques we use. 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 (9) and (10) below. The two-level simulation point estimator of $\mathrm{ES}_{1-p}$ is

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

where $\pi_1$ is a permutation of $\{1, 2, \ldots, k\}$ such that $\bar{X}_{\pi_1(1)}(N_{\pi_1(1)}) \leqslant \bar{X}_{\pi_1(2)}(N_{\pi_1(2)}) \leqslant \cdots \leqslant \bar{X}_{\pi_1(k)}(N_{\pi_1(i)})$. If $N_i = 0$, then $\bar{X}_i$ is taken to be $\infty$ 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 $\alpha$ can be decomposed as $\alpha = \alpha_o + \alpha_i$, representing errors due to the outer and inner levels of simulation, respectively. Here we further decompose $\alpha_i = \alpha_s + \alpha_{hi} + \alpha_{lo}$, where $\alpha_s$ is error due to screening and $\alpha_{hi}$ and $\alpha_{lo}$ are errors, respectively, associated with upper and lower confidence limits for inner-level simulation.

Before providing a detailed description of our procedure and how it produces an asymptotically valid confidence interval from a two-level simulation, we explain why we advocate using such a complicated procedure to produce a confidence interval. Looking at Equation (3), one might think to run $m$ separate two-level simulations and use the resulting point estimates $\widehat{\mathrm{ES}}_{1-p}^{(1)}, \widehat{\mathrm{ES}}_{1-p}^{(2)}, \ldots, \widehat{\mathrm{ES}}_{1-p}^{(m)}$ to construct a confidence interval for $\mathrm{ES}_{1-p}$, e.g., based on the assumption that these point estimates are normally distributed with mean $\mathrm{ES}_{1-p}$. However, the point estimator is biased because of inner-level sampling uncertainty (Lan et al. 2007b). Consequently, in experiments not reported here, we found that the normal-theory confidence interval had very low coverage except when it was so wide as to be useless for practical purposes. Our procedure takes the bias into account when constructing the confidence interval and thus is able to provide a confidence interval that is narrow and achieves the nominal coverage.

### 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 $\Pr\{\gamma \subseteq I\} \geqslant 1 - \alpha_s$. The number of pairwise comparisons between $\gamma$ 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 $V_i \leqslant V_j$ 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 $S_{ij}^2(n_0)$ be the sample variance of $X_{i1} - X_{j1}$, $X_{i2} - X_{j2}, \ldots, X_{in_0} - X_{jn_0}$. We retain all scenarios that are beaten fewer than $\lceil kp \rceil$ times:

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

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. Because the $\lceil kp \rceil$ scenarios with the lowest sample averages are necessarily

beaten fewer than $\lceil kp \rceil$ times, they survive screening, so $|I| \geqslant \lceil kp \rceil$. Because of the extra margin $dS_{ij}(n_0)/\sqrt{n_0}$ in Equation (4), there are pairs $(i, j)$ such that neither does scenario $i$ beat scenario $j$, nor does scenario $j$ beat scenario $i$; therefore, it is possible for $|I| > \lceil kp \rceil$ scenarios to survive screening.

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 $X_i - X_j$ by inducing positive correlation between $X_i$ and $X_j$. In finance, CRN usually induces a substantial positive correlation between $X_i$ and $X_j$. In our examples, $X_i$ and $X_j$ have to do with stock option payoffs at maturity $U > T$ simulated conditional on the stock price at time $T$ being $S_i(T)$ or $S_j(T)$, respectively. These payoffs are highly correlated when CRN is used, because, in our examples, CRN makes the stock return between times $T$ and $U$ the same for any value of $S(T)$.

### 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 $\mathrm{ES}_{1-p}$ based on empirical likelihood is discussed by Baysal and Staum (2008). Here we review a few essential facts for understanding the operation of empirical likelihood in our two-level simulation procedure.

Empirical likelihood involves considering distributions that arise by 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

$$\mathscr{S}(k) := \bigcup_{l=1}^{k} \mathscr{S}_l(k),$$

where

$$\mathscr{S}_l(k) := \left\{ \mathbf{w}\colon \mathbf{w} \geqslant 0, \sum_{i=1}^{k} w_i = 1, \sum_{i=1}^{l} w_i = p, \prod_{i=1}^{k} w_i \geqslant c\, k^{-k} \right\}, \quad (6)$$

where $c$ is a critical value derived from a chi-squared distribution. Each $\mathbf{w} \in \mathscr{S}(k)$ belongs to $\mathscr{S}_l(k)$ for a unique integer $l$, which can be interpreted as the number of scenarios that we believe belong to the tail of $F_V$, i.e., are less than or equal to the $p$-quantile of $V$. The intuition behind using only weight vectors that fall in $\mathscr{S}(k)$ in this application of empirical likelihood is as follows. First, it is unlikely that too few or too many of the scenarios that we sampled from $F_V$ belong to the tail of $F_V$. There are integers $l_{\min}$ and $l_{\max}$ such that $\mathscr{S}_l(k)$ is empty if $l < l_{\min}$ or $l > l_{\max}$; we need only consider a limited range of $l$, not all $1, 2, \ldots, k$. (Although $l$ depends on $\mathbf{w}$, and $l_{\min}$ and $l_{\max}$ depend on $k$, to lighten notation we do not make this dependence explicit.) Second, we use only weight vectors that are fairly close to uniform; if elements $w_i$ are too far from $1/k$, then $\prod_{i=1}^{k} w_i < c\, k^{-k}$. This means that we work with discrete distributions that are not too far from the empirical distribution, which places weight $1/k$ on each of $V_1, V_2, \ldots, V_k$.

Because $\mathrm{ES}_{1-p}$ involves an average over the left tail containing probability $p$, we also define a transformed weight vector $\mathbf{w}'$:

$$w_i' := \begin{cases} -w_i/p, & i = 1, 2, \ldots, l \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

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 $w_i$ to each value $V_i$. For this distribution, $\mathrm{ES}_{1-p}$ is $\sum_{i=1}^{k} w_i' V_{\pi_V(i)}$. The empirical likelihood confidence interval for $\mathrm{ES}_{1-p}$ of the unknown true distribution $F_V$, expressed in Equation (1), is

$$\left[ \min_{\mathbf{w} \in \mathscr{S}(k)} \sum_{i=1}^{k} w_i' V_{\pi_V(i)},\ \max_{\mathbf{w} \in \mathscr{S}(k)} \sum_{i=1}^{k} w_i' V_{\pi_V(i)} \right], \quad (8)$$

representing the outer-level uncertainty entailed by working with a sample $Z_1, Z_2, \ldots, Z_k$ instead of the true distribution $F_Z$. The intuition behind (8) is that $\mathscr{S}(k)$ is the set of weight vectors that are "empirically likely," and we do not believe that the true ES according to $F_V$ is less than the smallest ES that comes from applying weights in $\mathscr{S}(k)$ to the values $\mathbf{V}$, nor do we believe that the true ES is more than the largest ES that comes from applying weights in $\mathscr{S}(k)$ to the values $\mathbf{V}$.

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 the electronic companion, is that the lower confidence limit is

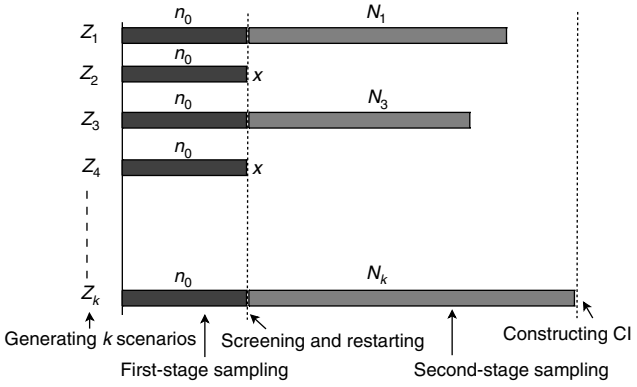
$$\min_{l=\lfloor kp \rfloor, \ldots, l_{\max}} \left( \min_{\mathbf{w} \in \mathscr{S}_l(k)} \sum_{i=1}^{l} w_i' \bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - t_{lo}(l) B_0(l) \right) \quad (9)$$

and the upper confidence limit is

$$\max_{l=l_{\min}, \ldots, \lceil kp \rceil} \left( \max_{\mathbf{w} \in \mathscr{S}_l(k)} \sum_{i=1}^{l} w_i' \bar{X}_{\pi_1(i)}(N_{\pi_1(i)}) + t_{hi} B_S(l) \right), \quad (10)$$

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 (9) arises from that in Equation (8) 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 (10) 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 $v_p$.

**Figure 1.** Schematic illustration of our procedure's operation.




### 3.3. Outline of 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 is 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, illustrated in Figure 1:

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 $\pi_0$ 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) *Screening of scenario* $\pi_0(i)$*:* 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 $b \geqslant \lceil kp \rceil$, scenario $\pi_0(i)$ is screened out: go to Step 3(b). Otherwise, set $j \leftarrow j + 1$.

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

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

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

$$N_i := \left\lceil \frac{C_1 S_i^2(n_0)}{\sum_{j \in I} S_j^2(n_0)} \right\rceil. \tag{11}$$

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

$$l_{\min} := \min\left\{ l : k^k \left(\frac{p}{l}\right)^l \left(\frac{1-p}{k-l}\right)^{k-l} \geqslant c \right\} \quad \text{and}$$

$$l_{\max} := \max\left\{ l : k^k \left(\frac{p}{l}\right)^l \left(\frac{1-p}{k-l}\right)^{k-l} \geqslant c \right\},$$

respectively, the smallest and largest numbers of scenarios to use in estimating ES.

Initialize the lower confidence limit $\hat{L} \leftarrow \infty$. Compute $N_{lo}(\lceil kp \rceil) := \min_{i=1,2,\ldots,\lceil kp \rceil} N_{\pi_0(i)}$ and $\underline{s}(\lceil kp \rceil) := \max_{i=1,2,\ldots,\lceil kp \rceil} s_{\pi_0(i)}$, which are, respectively, the smallest sample size and the largest standard error associated with any of the $\lceil kp \rceil$ scenarios with the lowest first-stage sample averages.

The following loop computes the lower confidence limit as a minimum of lower bounds associated with different numbers $l$ of scenarios that could be used in estimating ES. For $l = \lfloor kp \rfloor, \lfloor kp \rfloor + 1, \ldots, l_{\max}$:

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

$$\Delta(l) := \sqrt{\max_{\mathbf{w} \in \mathcal{I}(l)} \sum_{i=1}^{l} (w_i')^2}, \tag{12}$$

and $B_0(l) := \underline{s}(l)\Delta(l)$, which serves to bound standard error in estimating ES using $l$ scenarios.

(b) Set

$$\hat{L} \leftarrow \min\left\{ \hat{L}, \min_{\mathbf{w} \in \mathcal{I}_l(k)} \sum_{i=1}^{l} w_i' \bar{X}_{\pi_0(i)}(N_{\pi_0(i)}) - t_{lo}(l) B_0(l) \right\}. \tag{13}$$

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

Sort to produce a mapping $\pi_1$ from $\{1, 2, \ldots, |I|\}$ to $I$ such that $\bar{X}_{\pi_1(i)}(N_{\pi_1(i)})$ is nondecreasing in $i$.

Initialize the upper confidence limit $\hat{U} \leftarrow -\infty$ and the largest standard error associated with any scenario $\bar{s} := \max_{i=1,2,\ldots,k} s_i$.

Compute the smallest sample size associated with any scenario that survived screening, $N_{hi} := \min\{N_{\pi_1(1)}, N_{\pi_1(2)}, \ldots, N_{\pi_1(|I|)}\}$, and $t_{hi} := t_{1-\alpha_{hi}, N_{hi}-1}$.

The following loop computes the upper confidence limit as a maximum of upper bounds associated with different numbers $l$ of scenarios that could be used in estimating ES. For $l = l_{\min}, l_{\min}+1, \ldots, \lceil kp \rceil$:

(a) Compute $\Delta(l)$ as in Equation (12) and $B_S(l) := \bar{s}\Delta(l)$ to bound standard error in estimating ES using $l$ scenarios.

(b) Set

$$\hat{U} \leftarrow \max\left\{\hat{U}, \max_{\mathbf{w}\in\mathscr{S}_l(k)} \sum_{i=1}^{l} w_i' \bar{X}_{\pi_1(i)}(N_{\pi_1(i)}) + t_{hi} B_S(l)\right\}. \quad (14)$$

The confidence interval given in Equations (9) and (10) is $[\hat{L}, \hat{U}]$.

The maximum in Equation (12) is computed by a method described in the appendix. An algorithm for the optimizations in Equations (13) and (14) is given in Baysal and Staum (2008). The $t$-quantiles $t_{lo}(l)$ and $t_{hi}$ may be replaced by normal quantiles when the second-stage sample sizes are sufficiently large, as they typically are.

### 3.4. Sketch of the Proof of Asymptotic Validity of the Confidence Interval

Due to its length, the complete proof of the asymptotic validity of the confidence interval as the number of scenarios $k \to \infty$ appears in the electronic companion. Here we sketch the main ideas of the proof.

The basic approach is error spending: we decompose the allowable error probability $\alpha = \alpha_i + \alpha_o$, where $\alpha_i$ and $\alpha_o$ are the allowed probabilities of an error associated with inner- and outer-level simulation, respectively. This is a feature of the general framework of Lan et al. (2007b), showing how to construct an asymptotically valid confidence interval for two-level simulation from the following two components. The first is a confidence interval that is a function of the vector $\mathbf{V} := (V_1, V_2, \ldots, V_k)$ of unknown true portfolio values and whose coverage probability is at least $1-\alpha_o$ in the limit as the outer-level sample size $k \to \infty$. In our case, this is the outer-level confidence interval of Equation (8), and its asymptotic validity is proved by Baysal and Staum (2008). The second component we use is a confidence region $\mathscr{V}$ for the vector $\mathbf{V}$ whose coverage probability is $1-\alpha_i$.

Our proof shows that the confidence limits $L$ and $U$ in Equations (9) and (10) are indeed constructed from Equation (8) and a confidence region $\mathscr{V}$ in a way that fits the framework of Lan et al. (2007b), and that this confidence region $\mathscr{V}$ has coverage probability $1-\alpha_i$. To simplify our proof, we assume that the payoffs are normally distributed so as to work with a confidence region $\mathscr{V}$ that is valid even at finite inner-level sample size. (Because of the central limit theorem, we expect the payoffs to be approximately normal if the sample sizes $n_0$ and $N_i$ are large enough.) The proof uses a further error-spending decomposition $\alpha_i = \alpha_{hi} + \alpha_{lo} + \alpha_s$, in which the error probabilities are associated with a violation of the upper confidence limit in Equation (8), with a violation of the lower confidence limit in Equation (8), and with screening out a scenario that should not be screened out, respectively. We use the Bonferroni inequality to show that the error probability associated with screening is bounded by $\alpha_s$, in a way standard in the ranking-and-selection literature. The novelty is in the shape chosen for the confidence region $\mathscr{V}$ and the arguments that bound error probabilities by $\alpha_{hi}$ and $\alpha_{lo}$. The key to bounding error probabilities by $\alpha_{hi}$ and $\alpha_{lo}$ is Lemma EC.2, which provides a tool like a $t$-test for weighted sums of independent normal random variables.

## 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 §2. The error $\alpha = 10\%$ was decomposed into $\alpha_o = 5\%$ for the outer level, $\alpha_s = 2\%$ for screening, and $\alpha_{lo} = \alpha_{hi} = 1.5\%$ 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 (2010). 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 (9) and (10). 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 $k \geqslant 40/p$, 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 $ES_{0.99}$. 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 $ES_{0.99}$. For these reasons, we ran experiments with computational budgets such that our procedure yields CI widths in the neighborhood of 10% relative error.

Figure 2 shows how average CI width varies with a computational budget of $C$ replications for the example of selling a put option described in §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 2, 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 3 shows similar results from the example of an options portfolio described in § 2.2. In this example, larger computational budgets are required to get a 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 $k = 4{,}000$ and $n_0 = 4{,}703$, 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 multistage screening procedure (similar to Lesnevski et al. 2007, 2008) might overcome this problem.

**Figure 2.** Average confidence interval width in the example of §2.1 given a fixed budget of simulation replications.

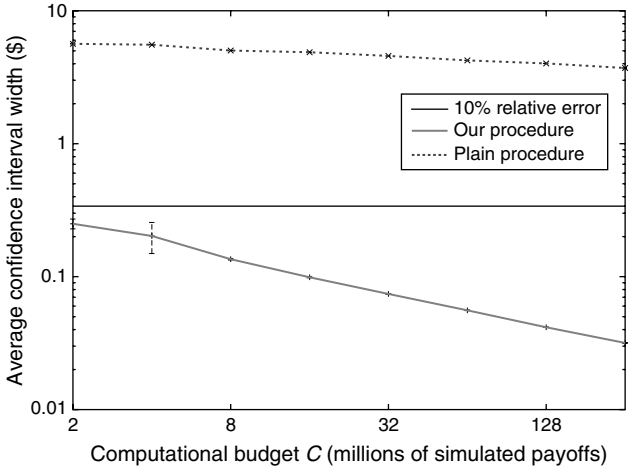
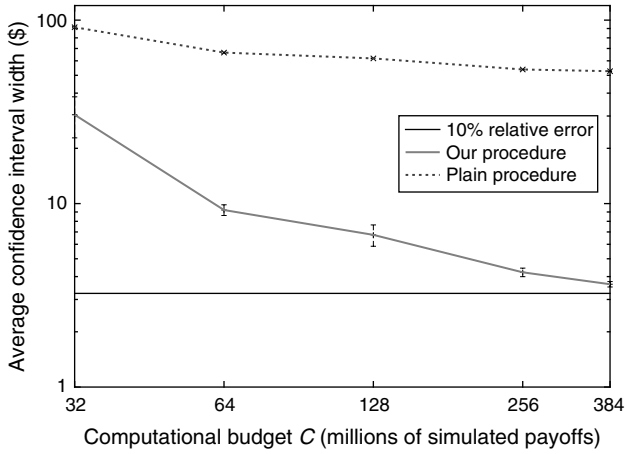
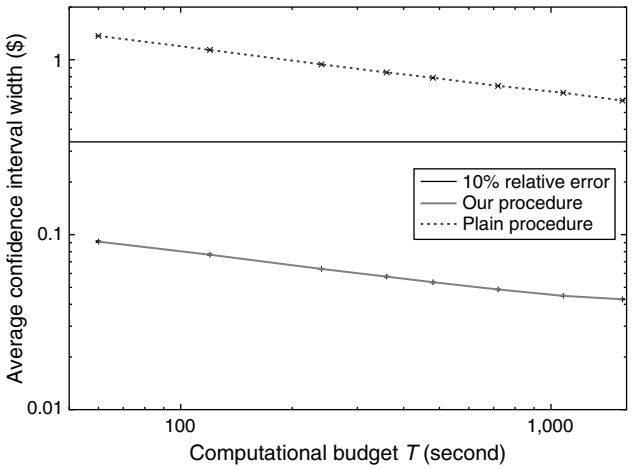



**Figure 3.** Average confidence interval width in the example of §2.2 given a fixed budget of simulation replications.




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 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 2010). For instance, in the example of a single put option (§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 4, where the curve representing our procedure's CI width becomes flatter as the computing time $T$ grows. Still, Figures 4 and 5 show that our procedure performs much better than the plain procedure when they

**Figure 4.** Average confidence interval width in the example of §2.1 given a fixed budget of computing time.



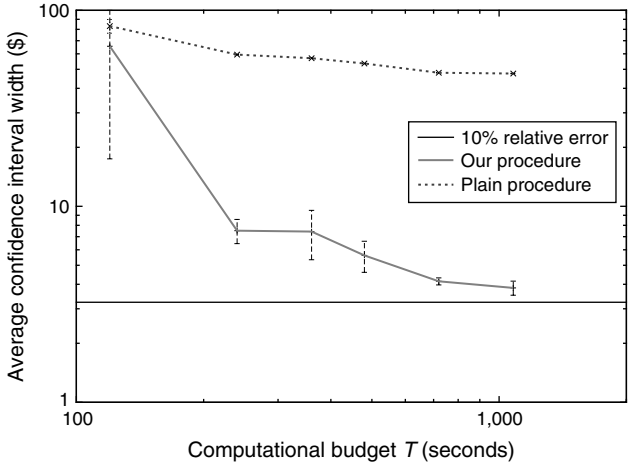

are given equal computing times, producing a CI narrower by a factor of as much as 15 or 12 in these two examples.

## 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 $k \geqslant 40/p$, 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) mention potential enhancements to the empirical likelihood estimation used here. As in Lesnevski et al. (2008), it may help to use multistage 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 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.

**Figure 5.** Average confidence interval width in the example of §2.2 given a fixed budget of computing time.




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 $\alpha$ 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 $\alpha$ in a way that makes the procedure efficient: the values we chose in §4 are broadly effective, a finding that 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 multistage 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 (2010).

## 6. Electronic Companion

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

## Appendix. Computation of the Maximum in Equation (12)

By the definition of $\mathbf{w}'$ in Equation (7), the computation of $\Delta(l) := \sqrt{\max_{\mathbf{w}\in\mathcal{S}(l)} \sum_{i=1}^{l} (w_i')^2}$ in Equation (12) is equivalent to maximizing $\sum_{i=1}^{l} w_i^2/p^2$ over the set $\mathcal{S}(l) := \{\mathbf{w} : \mathbf{w} \geqslant 0,\ \sum_{i=1}^{k} w_i = 1,\ \sum_{i=1}^{l} w_i = p,\ \prod_{i=1}^{k} w_i \geqslant c\,k^{-k}\}$. Choosing $w_{l+1} = w_{l+2} = \cdots = w_k = (1-p)/(k-l)$ has no effect on the objective and leads to the loosest possible constraint on $w_1, w_2, \ldots, w_l$, namely, $\prod_{i=1}^{l} w_i \geqslant ck^{-k}((1-p)/$

$(k-l))^{-(k-l)}$. Thus, letting $x_i = w_i/p$ for $i = 1, 2, \ldots, l$, we have reduced the problem to the $l$-dimensional problem

$$\max f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{x} \in \mathscr{X} := \{\mathbf{x}\colon \mathbf{x} \geqslant 0, g(\mathbf{x}) \geqslant 0, h(\mathbf{x}) = 0\},$$

where $f(\mathbf{x}) = \sum_{i=1}^{l} x_i^2$, $g(\mathbf{x}) = \sum_{i=1}^{l} \log x_i + l \log p - \log c + k \log k - (k-l)(\log(k-l) + \log(1-p))$ and $h(\mathbf{x}) = \sum_{i=1}^{l} x_i - 1$. The function $f$ is convex, and the feasible set $\mathscr{X}$ is nonempty, closed, bounded, and convex. Therefore, the maximum is attained at an extreme point of $\mathscr{X}$ (Rockafellar 1970, Corollary 32.3.2).

We next turn to necessary conditions on the gradients $\nabla f$, $\nabla g$, and $\nabla h$ at any point $\mathbf{x}^*$ where the maximum is attained. The gradients of the objective and of the active constraints are given by $(\nabla f(\mathbf{x}))_i = 2x_i$, $(\nabla g(\mathbf{x}))_i = 1/x_i$, and $(\nabla h(\mathbf{x}))_i = 1$. The constraint $\mathbf{x} \geqslant 0$ is not active at any $\mathbf{x} \in \mathscr{X}$ because having $x_i = 0$ for any $i$ leads to a violation of the constraint $g(\mathbf{x}) \leqslant 0$. Because $\mathbf{x}^*$ is an extreme point of $\mathscr{X}$, its coordinates are not all equal, so the rank of $[\nabla g(\mathbf{x}) \nabla h(\mathbf{x})]$ is two. Because this equals the number of active constraints, there exist $\lambda \geqslant 0$ and $\mu \geqslant 0$ such that $\nabla f(\mathbf{x}^*) + \lambda \nabla g(\mathbf{x}^*) + \mu \nabla h(\mathbf{x}^*) = 0$ (Sundaram 1996, Theorem 6.10). That is, $2x_i^* + \lambda/x_i^* + \mu = 0$ for all $i = 1, 2, \ldots, l$. This equation has at most two positive real roots; therefore, any point $\mathbf{x}^*$ where the maximum is attained is an extreme point of $\mathscr{X}$ with the property that the set $\{x_1^*, x_2^*, \ldots, x_l^*\}$ contains at most two distinct values.

The only remaining questions are what those two values are and how many of the coordinates $x_1^*, x_2^*, \ldots, x_l^*$ take on each of the two values—by symmetry, it does not matter how the coordinates are permuted. Consider an extreme point $\mathbf{x}^{(m)}$ of $\mathscr{X}$ such that $m$ coordinates take on one value, $a^{(m)}$, whereas the other $l - m$ coordinates take on another value, $b^{(m)}$. Because the order does not matter, we take $x_1^{(m)} = x_2^{(m)} = \cdots = x_m^{(m)} = a^{(m)}$ whereas $x_{m+1}^{(m)} = x_{m+2}^{(m)} = \cdots x_l^{(m)} = b^{(m)}$. To satisfy the constraint $h(\mathbf{x}^{(m)}) = 0$, we must have $b^{(m)} = (p - ma^{(m)})/(l - m)$. For each $m = 1, 2, \ldots, l-1$, we compute $a^{(m)}$ by solving $g(\mathbf{x}^{(m)}) = 0$. Finally, we compute $\Delta(l) = \max\{f(\mathbf{x}^{(m)})\colon m = 1, 2, \ldots, l-1\}$.

## Acknowledgments

This article is based on work supported by the National Science Foundation under grant DMI-0555485. The authors are grateful to the associate editor and referees for corrections and comments that led to improvements in the article.

## References

Acerbi, C., D. Tasche. 2002. On the coherence of expected shortfall. *J. Banking and Finance* **26**(7) 1487–1503.

Baysal, R. E., J. Staum. 2008. Empirical likelihood for value at risk and expected shortfall. *J. Risk* **11**(1) 3–32.

Boesel, J., B. L. Nelson, S.-H. Kim. 2003. Using ranking and selection to "clean up" after simulation optimization. *Oper. Res.* **51**(5) 814–825.

Chen, S. 2008. Nonparametric estimation of expected shortfall. *J. 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. B., S. Juneja. 2006. Efficient simulation for risk measurement in portfolio of CDOs. L. F. Perrone, B. Lawson, J. Liu, F. P. Wieland, eds. *Proc. 2006 Winter Simulation Conf.* IEEE Press, Piscataway, NJ, 749–756.

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

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

Lan, H., B. L. Nelson, J. Staum. 2007a. A confidence interval for tail conditional expectation via two-level simulation. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, R. R. Barton, eds. *Proc. 2007 Winter Simulation Conf.* IEEE Press, Piscataway, NJ, 949–957.

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

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

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

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

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

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

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

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

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.

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

Sundaram, R. K. 1996. *A First Course in Optimization Theory*. Cambridge University Press, New York.