

# VARIANCE OF THE CTE ESTIMATOR

B. John Manistre\* and Geoffrey H. Hancock†

---

## ABSTRACT

The Conditional Tail Expectation (*CTE*), also called Expected Shortfall or Tail-*VaR*, is a robust, convenient, practical, and coherent measure for quantifying financial risk exposure. The *CTE* is quickly becoming the preferred measure for statutory balance sheet valuation whenever real-world stochastic methods are used to set liability provisions. We look at some statistical properties of the methods that are commonly used to estimate the *CTE* and develop a simple formula for the variance of the *CTE* estimator that is valid in the large sample limit. We also show that the formula works well for finite sample sizes. Formula results are compared with sample values from real-world Monte Carlo simulations for some common loss distributions, including equity-linked annuities with investment guarantees, whole life insurance and operational risks. We develop the *CTE* variance formula in the general case using a system of biased weights and explore importance sampling, a form of variance reduction, as a way to improve the quality of the estimators for a given sample size. The paper closes with a discussion of practical applications.

---

## 1. INTRODUCTION

The Conditional Tail Expectation, also called Expected Shortfall or Tail Value-at-Risk, is a robust, convenient and coherent measure for quantifying risk exposure. For example, the *CTE* is the preferred measure for establishing the Canadian statutory balance sheet provisions on equity-linked variable insurance and annuity products with investment guarantees. This paper examines some statistical properties of the methods that are commonly used to estimate the *CTE* and compares theoretical results to sample values from real-world simulations.

While the *CTE* is relatively new to actuarial practice,<sup>1</sup> there are good reasons for believing that it is here to stay and will grow, rather than diminish, in importance. First, the *CTE* is a simple, intuitive, and useful example of a tail risk measure and such measures are now widely used in many areas of financial analysis. Second, the *CTE* is coherent in the sense of Artzner (1999) and less sensitive to sampling error than the more traditional quantile or Value-at-Risk (*VaR*) measures. Given the importance of this measure, especially for the assignment of solvency capital for insurance enterprises, it makes sense to fully understand its properties. This paper contributes to that understanding.<sup>2</sup> Our primary audience is the practicing actuary who uses the *CTE* as a tool to set balance sheet provisions and manage risk exposure.

Following this introduction, this paper is organized into five main sections. In Section 2, we outline the theory that supports the formula for the variance of the *CTE* estimator in the large sample limit. Section 3 discusses the consequences of using the formula for finite samples by reference to the Pareto

---

\* B. John Manistre FSA, FCIA, MAAA, is a Vice President, Corporate Actuarial at Aegon USA Inc., 4333 Edgewood Road NE, Cedar Rapids, IA 52499, jmanistre@aegonusa.com.

† Geoffrey H. Hancock FSA, FCIA, is a Director of Mercer Oliver Wyman, BCE Place at Mercer Oliver Wyman, BCE Place, 161 Bay Street, PO Box 501, Toronto, ON M5J 2S5, Canada, ghancock@mow.com.

<sup>1</sup> The *CTE* became the de facto standard measure for Canadian statutory valuation of balance sheet liabilities in 2001 and is recommended by the American Academy of Actuaries in its proposal for setting risk-based capital for variable annuities. See CIA Task Force (2002), CIA Committee on Life Insurance Financial Reporting (2001), and AAA Life Capital Adequacy Subcommittee (2002).

<sup>2</sup> This paper builds on the work of Manistre and Hancock (2002). See Jones and Zitikis (2003) for similar results related to risk measures other than the *CTE*.

distribution. Section 4 tests the variance formula with real-world examples obtained by Monte Carlo simulation. Section 5 addresses the issue of sampling error and ways of reducing it with variance reduction. Finally, the paper concludes with a discussion of the practical implications of using the *CTE* measure in the insurance industry and how understanding its variance can contribute to better risk management.

A more precise definition of the problem under discussion is now provided. Suppose we want to estimate the Conditional Tail Expectation of a random variable  $X$ , with cumulative distribution  $\Pr\{X \leq x\} = F(x)$  at the level  $\alpha$ . Thus, we want to calculate the conditional expectation

$$CTE(\alpha) = E[X|X > q_\alpha]$$

where  $q_\alpha$  is the  $\alpha$ -quantile, defined as the smallest value satisfying

$$\Pr\{X > q_\alpha\} = 1 - \alpha.$$

The  $\alpha$ -quantile is often called Value-at-Risk<sup>3</sup> (*VaR*) and is used extensively in the financial management of trading risk over a fixed (usually short) time horizon.

Note, however, that our *CTE* definition is ambiguous when  $q_\alpha$  falls in a probability mass. In this case, we need a more precise definition such as that provided by Hardy (2001). However, without loss of generality and to simplify the notation, we will ignore this distinction in the remainder of the paper.

When  $X$  is unknown, the standard approach to this problem is to start with a random sample  $(x_1, x_2, \dots, x_n)$  of size  $n$  from the distribution  $F(x)$  and then sort the sample in descending order to obtain the *order statistics*<sup>4</sup>  $(x_{(1)} \geq x_{(2)} \geq \dots \geq x_{(n)})$ . Given these order statistics, the *CTE* estimator at the  $\alpha = 1 - k/n$  level is given by the average of the  $k$  highest order statistics:

$$\hat{CTE}(\alpha) = \frac{1}{k} \sum_{j=1}^k x_{(j)}.$$

It is not hard to show that the *CTE* estimator is *asymptotically* unbiased and coherent, which makes it a superior risk measure to *VaR*.<sup>5</sup> The variance of the estimator is harder to obtain because the order statistics are not independent. This means that the obvious quantity for the standard deviation of the *CTE* estimator (*SDE*) is not a valid formula:

$$SDE = \sqrt{\frac{1}{k} \sum_{j=1}^k \frac{(x_{(j)} - \hat{CTE})^2}{(k-1)}}.$$

Empirical studies show that *SDE* can materially understate the true uncertainty in the estimator. This can be understood at a high level by noting that in calculating the empirical *CTE* we are actually estimating two quantities. These are:

1. The  $\alpha = 1 - k/n$  quantile. We are using the  $k^{\text{th}}$  sample order statistic  $x_{(k)}$  for this.
2. The conditional expectation of  $X$ , given that it exceeds the  $\alpha$ -quantile.

Uncertainty in both quantities contributes to the uncertainty in the *CTE* estimate, which is why the quantity *SDE* defined above understates the true variance. Another way to see this is to condition on the observation of the  $k^{\text{th}}$  order statistic and write

<sup>3</sup> As discussed in Artzner (1999), some common measures such as *VaR* and semivariance do not satisfy the four axioms for coherence. However, the *CTE* is a coherent risk measure as defined therein. By itself, coherence does not guarantee the suitability of a risk measure as a management device in all situations, but violation of the axioms can indicate significant flaws and potential for failure.

<sup>4</sup> The order statistics are usually defined as  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ , but it will be convenient to use our definition in this paper.

<sup>5</sup> Wang (2002) suggests that the *CTE* lacks incentive for mitigating losses below the quantile "*VaR*" and furthermore does not properly adjust for extreme low-frequency and high-severity losses since it only accounts for expected shortfall. He argues for a family of coherent risk measures based on distorted probabilities.

$$VAR(C\hat{T}E) = E [VAR(C\hat{T}E|X_{(k)})] + VAR[E(C\hat{T}E|X_{(k)})].$$

The quantity  $1/k \cdot VAR(x_{(1)}, x_{(2)}, \dots, x_{(k)})$  is an unbiased estimate of the first term above, but it does not capture the second term. It turns out that the second term can be estimated in the large sample limit (as  $n \rightarrow \infty$ ). The main result of this paper, which deals with both sources of uncertainty, is:

$$VAR(C\hat{T}E) \approx \frac{VAR(x_{(1)}, \dots, x_{(k)}) + \alpha \cdot (C\hat{T}E - x_{(k)})^2}{k}.$$

This formula has much intuitive appeal as it gives the well-known variance of a mean estimator  $\sigma^2/n$  when  $\alpha = 0$ . This agrees with our definition for *CTE* since  $C\hat{T}E(0)$  is the unconditional sample mean  $\bar{x}$ .

Interestingly, the second term gains importance when the probability distribution is light-tailed. An example of a light-tailed risk is an “in-the-money” investment guarantee on an equity-linked variable insurance product. When the distribution is heavy-tailed, the first term for the estimator variance will dominate.

## 2. UNDERLYING THEORY

We now outline the theory that supports the variance estimator in the large sample limit. Literature references are provided for those who wish to pursue the theory in more detail. Readers who are willing to accept these theoretical results can skip the first part of this section without loss of continuity. The example of a simple European put option developed at the end of this section will be used again to illustrate some variance reduction techniques in Section 5.

We develop results for the joint distribution of the *CTE* and quantile estimators for an unbiased sampling method. Biased sampling methods to reduce variance are discussed in Section 5.

Parameter estimation is a highly developed statistical subject. One of the tools that has been developed by statisticians to assess the errors in parameter estimation is a process called the *delta method*.<sup>6</sup> The delta method is well known in the statistical community, but it is not a standard topic in actuarial study. In its essence, the delta method expands a function of a random variable about its mean, usually with a Taylor approximation, and then takes the variance.

This section is designed to describe the key theoretical results that follow from the delta method and are used in this paper. More precise statements and detailed proofs can be found in the cited literature.

The delta method applies when the parameters  $\theta$  being estimated can be represented as functionals  $\theta = T(F)$  of an underlying probability distribution  $F$ . If the parameters are estimated by applying the functional  $T$  to an empirical distribution  $\hat{F}_n$ , derived from a random sample of size  $n$ , i.e.,  $\hat{\theta}_n = T(\hat{F}_n)$ , then the delta method asserts that, under certain technical conditions, the parameter estimates are asymptotically unbiased and normally distributed.

For the remainder of this paper, we use *VaR* to denote the Value-at-Risk estimator ( $\alpha$ -quantile), and *VAR* to represent variance.

The *CTE* can be represented by a Stieltjes integral<sup>7</sup> of the form

$$CTE(\alpha) = T_\alpha(F) = \int x d\hat{g}_\alpha[F(x)]$$

where  $\hat{g}_\alpha : [0,1] \rightarrow [0,1]$  is the continuous function given by

<sup>6</sup> The theory summarized in this section can be found in Staudte and Sheather (1990). Most of the key ideas can be found in section 3.2 and Appendix B. The authors are grateful to Dr. Jacques Rioux, ASA for this reference.

<sup>7</sup> The Stieltjes integral  $\int f dg$  is defined by taking limits of the form  $\sum f(x_i)[g(x_{i+1}) - g(x_i)]$ . The limit is well defined if  $f$  is continuous and  $g$  is monotone increasing.

$$g_{\alpha}(t) = \begin{cases} 0 & t < \alpha \\ \frac{t - \alpha}{1 - \alpha} & t \geq \alpha \end{cases}.$$

Similarly, the quantile or Value-at-Risk can be represented as a functional of the form

$$VaR(\alpha) = H_{\alpha}(F) = \int x dh_{\alpha}[F(x)]$$

where  $h_{\alpha} : [0,1] \rightarrow [0,1]$  is given by

$$h_{\alpha}(t) = \begin{cases} 0 & t < \alpha \\ 1 & t \geq \alpha \end{cases}.$$

When a random sample  $(x_1, \dots, x_n)$  of size  $n$  is drawn from the distribution  $F$ , we can form an empirical distribution  $\hat{F}_n(x)$  by letting

$$\hat{F}_n(x) = \frac{1}{n} \sum_i \begin{cases} 1 & \text{if } x_i \leq x \\ 0 & \text{if } x_i > x \end{cases}$$

and we can then calculate the *plug-in estimators* for the required parameters by applying the appropriate functional to the empirical distribution

$$CT\hat{E}_n(\alpha) = \int x dg_{\alpha}[\hat{F}_n(x)]$$

$$V\hat{a}R_n(\alpha) = \int x dh_{\alpha}[\hat{F}_n(x)].$$

It can be verified that the plug-in estimators for the *CTE* and *VaR* agree with the estimators described earlier provided we choose  $k$  to be the largest integer such that  $1 - k/n > \alpha$ . More practical expressions for the plug-in estimators are clearly

$$CT\hat{E}_n(\alpha) = \frac{1}{k} \sum_{i=1}^k x_{(i)},$$

$$V\hat{a}R_n(\alpha) = x_{(k)}.$$

Now let  $\Delta_y(x)$  be the distribution function of a point mass concentrated at  $x = y$  and consider the distribution function obtained by interpolating between  $F(x)$  and  $\Delta_y(x)$

$$F_{y,\varepsilon}(x) = (1 - \varepsilon) \cdot F(x) + \varepsilon \cdot \Delta_y(x).$$

The *Influence Function* of the parameter  $\theta$  is then defined by the limit

$$IF_{\theta}(y) = \lim_{\varepsilon \rightarrow 0} \frac{T(F_{y,\varepsilon}) - T(F)}{\varepsilon},$$

if the limit exists. The Influence Function measures the relative influence on  $T(F)$  of a “bad” observation at  $y$  and plays the following role in the completion of the delta method.

Under certain technical conditions, the plug-in estimators are asymptotically unbiased and have a joint normal distribution.<sup>8</sup> The variance of a plug-in estimator satisfies

$$\lim_{n \rightarrow \infty} \left\{ \text{VAR}(\hat{\theta}_n) - \frac{1}{n} \cdot E[IF_{\theta}^2(X)] \right\} = 0,$$

and the covariance of two plug-in estimators  $\hat{\theta}_n, \hat{\lambda}_n$  satisfies

$$\lim_{n \rightarrow \infty} \left\{ \text{Cov}(\hat{\theta}_n, \hat{\lambda}_n) - \frac{1}{n} \cdot E[IF_{\theta}(X) \cdot IF_{\lambda}(X)] \right\} = 0.$$

<sup>8</sup> See footnote 6.

In order to apply this result, we need to know the Influence Functions of both the *CTE* and *VaR* parameters. The calculation of the influence function for *VaR* is given in Staudte and Sheather (1990). The result assumes that  $F$  has a well-defined density function  $f$  that is continuous and positive at the quantile

$$IF_{VaR}(x) = \begin{cases} \frac{-(1 - \alpha)}{f(VaR)} & x < VaR \\ 0 & x = VaR. \\ \frac{\alpha}{f(VaR)} & x > VaR \end{cases}$$

For the *CTE*, the method outlined by Staudte and Sheather has been used by J. Rioux<sup>9</sup> to obtain

$$IF_{CTE}(x) = \begin{cases} VaR - CTE & x < VaR \\ VaR - CTE + \frac{x - VaR}{1 - \alpha} & x > VaR \end{cases}$$

The asymptotic expressions for the variance of these estimators can now be calculated as follows. Here, we use the shortened notation  $IF_{\theta}$  to represent  $IF_{\theta}(X)$ , where  $X$  is the random variable and  $\theta$  is the estimator.

$$\begin{aligned} VAR(\hat{CTE}_n) &\approx \frac{E[(IF_{CTE})^2]}{n} = \frac{VAR(X|X > VaR) + \alpha \cdot (CTE - VaR)^2}{n \cdot (1 - \alpha)}, \\ VAR(\hat{VaR}_n) &\approx \frac{E[(IF_{VaR})^2]}{n} = \frac{\alpha \cdot (1 - \alpha)}{n \cdot [f(VaR)]^2}, \\ Cov(\hat{CTE}_n, \hat{VaR}_n) &\approx \frac{E[IF_{CTE} \cdot IF_{VaR}]}{n} = \frac{\alpha \cdot (CTE - VaR)}{n \cdot f(VaR)}. \end{aligned}$$

The first formula above is the asymptotic expression of interest. This is strong theoretical support for the use of this formula, but it does not guarantee that it will always work in practice.

The second formula shows that the variance of the quantile estimator depends on the value of the probability density of the underlying distribution at the quantile point  $f(VaR)$ . Estimating this quantity from data can be problematic, which is another practical reason for preferring the *CTE* to the *VaR* as a measure of risk.

We also see that the *VaR* and *CTE* estimators are positively correlated since  $CTE \geq VaR$  for a given  $\alpha$ -level. This makes intuitive sense because a random sample that leads to a positive error in the *VaR* will also overstate the *CTE*. We will see empirical evidence of this positive correlation later in our examples.

We close this section with a simple example to illustrate the theory developed above. Suppose we want to estimate the *CTE* of an “in-the-money” European put option<sup>10</sup> at the  $\alpha = 0.95$  confidence level by Monte Carlo simulation. To be more specific, assume the option matures in  $T = 10$  years with a strike price of  $X = 110$ . The current stock price is  $S = 100$  and assumed to follow a lognormal return process with  $\mu = 8\%$  and  $\sigma = 15\%$ . That is, the stock price at maturity is given by:

$$S(T) = S \cdot e^{[\mu T + \sigma \sqrt{T} Z]}$$

where  $Z$  is a standard Normal variate with mean zero and unit variance.

Using a continuous discount rate of  $\delta = 6\%$ , the random variable whose *CTE* we wish to calculate is then the *present value payoff function*:

<sup>9</sup> Private communication.

<sup>10</sup> A *European* put option gives the holder the right to sell the underlying asset on the maturity date for the specified strike price.

$$C = e^{-\delta T} \cdot \max[0, X - S \cdot e^{(\mu T + \sigma \sqrt{T} Z)}].$$

Using spreadsheet software, it is easy to generate  $n = 1000$  samples of this variable. From this sample, we can calculate the plug-in estimators for the CTE and VaR using the formulas developed earlier in this paper. To estimate the probability density  $f(VaR)$  we use the estimator:

$$\hat{f}(VaR) = \frac{\xi}{\hat{F}_n^{-1}(\alpha) - \hat{F}_n^{-1}(\alpha - \xi)}$$

with  $\xi = \frac{1}{100}$ . It is important to recognize that the empirical density function  $\hat{f}(\cdot)$  is quite sensitive to the choice of  $\xi$ , especially for small samples, even when the cumulative distribution function is “smooth.”

We can then calculate the Formula Standard Error (FSE) of each estimator as

$$FSE(CTE) = \sqrt{\frac{VAR(X_{(1)}, \dots, X_{(k)}) + \alpha \cdot (CTE - X_{(k)})^2}{n \cdot (1 - \alpha)}}$$

$$FSE(VaR) = \frac{1}{\hat{f}(VaR)} \cdot \sqrt{\frac{\alpha \cdot (1 - \alpha)}{n}}$$

$$C\hat{o}v(CTE, VaR) = \frac{\alpha \cdot (CTE - X_{(k)})}{n \cdot \hat{f}(VaR)}$$

Table 1 shows the results of two trials (first and last) and also the results of repeating the entire simulation 1000 times. The table also shows the exact values of the CTE and VaR for this problem, which can be calculated from the closed form expressions. Throughout this paper,  $\Phi(\cdot)$  represents the cumulative density function for the normal distribution with zero mean and unit variance.

$$VaR = e^{-\delta T} \cdot \max(0, X - S \cdot \exp[\mu T + \sigma \sqrt{T} \cdot z_\alpha]), \quad z_\alpha = \Phi^{-1}(1 - \alpha).$$

$$CTE = \frac{e^{-\delta T}}{1 - \alpha} \cdot \int_{-\infty}^{z_\alpha} \max(0, X - S \cdot \exp[\mu T + \sigma \sqrt{T} \cdot z]) \cdot \frac{\exp(-z^2/2)}{\sqrt{2\pi}} dz,$$

$$= \frac{e^{-\delta T}}{1 - \alpha} \cdot \left[ X \cdot \Phi(d_1) - S \cdot \Phi(d_1 - \sigma \sqrt{T}) \cdot e^{(\mu + \frac{1}{2}\sigma^2)T} \right], \quad d_1 = \min \left[ z_\alpha, \frac{\ln\left(\frac{X}{S}\right) - \mu \cdot T}{\sigma \sqrt{T}} \right].$$

Table 1 illustrates a number of significant points:

Table 1  
**Monte Carlo Simulation Without Variance Reduction**  
**CTE(95%) for a 10-year European Put Option (1000 Trials), X = \$110, S = \$100**

	$\hat{CTE}(95\%)$	$FSE(\hat{CTE})$	$\hat{VaR}$	$FSE(\hat{VaR})$	$C\hat{o}v(CTE, VaR)$	$\hat{f}(VaR)$
Closed Form	13.80	n/a	4.39	n/a	2.37	n/a
First Trial	13.67	1.54	5.09	1.40	1.65	0.0049
Last Trial	14.93	1.95	3.33	3.07	4.91	0.0022
Minimum	7.72	1.01	0	0.19	0.22	0.0009
Average	13.70	1.63	4.50	1.91	2.42	0.0040
Maximum	18.89	2.27	9.17	7.31	13.05	3.0065
Std Deviation	1.63	0.18	1.76	0.77	1.06	0.0019

- Any given trial provides a reasonable estimate of what would happen if the simulation were repeated, but with a sample size of  $n = 1000$  there is considerable variability, especially for  $\hat{VaR}$ .

- The *CTE* plug-in estimator is biased below the true closed form value of 13.80 (i.e., average  $C\hat{T}E$  is 13.70). This is always the case for small sample sizes.<sup>11</sup> However the bias is much smaller than the sampling error.
- The asymptotic variance formula for the *CTE* estimator performs quite well on average (i.e., average  $FSE(C\hat{T}E) = \text{empirical standard deviation of } C\hat{T}E$ ).
- In the formula for the sampling error of the *CTE* estimator, the second term is about 1.7 times larger than the first term. This is indicative of a light-tailed distribution and will be discussed in more detail in the next section.
- The *VaR* plug-in estimator is biased high (average is 4.50), but again the bias is much smaller than the sampling error.
- The sample covariance for all 1000 pairs of estimators is 2.37, which is higher (lower) than the estimated covariance from the first (last) trials, but close to the mean of all covariance estimators.

Figure 1  
**Monte Carlo Simulation Without Variance Reduction**  
**CTE(95%) for a 10-year European Put Option (10,000 Trials),  $X = \$110$ ,  $S = \$100$**

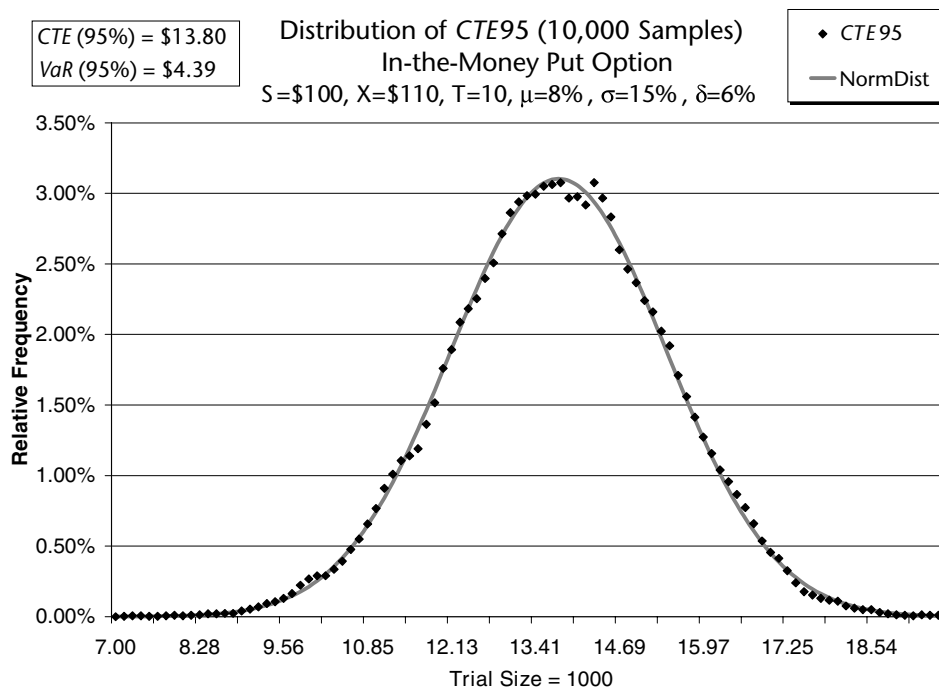


Figure 1 graphically depicts the empirical distribution of  $C\hat{T}E(95\%)$  using a sample size of 10,000. Due to the larger number of observations,<sup>12</sup> the statistics are slightly different from those presented in Table 1, but the conclusions remain unchanged. The relative error is defined as the standard deviation of the estimator divided by the mean. The column labelled “Normal” shows the values<sup>13</sup> for a normal distribution with mean and variance equal to the sample moments for  $C\hat{T}E$  and  $V\hat{a}R$  (as appli-

<sup>11</sup> In fact, the plug-in estimator is biased low for any finite sample size. To appreciate this, consider two samples of size  $N$ . The average of the two plug-in estimators must be less than the single estimate obtained when the two sets are combined to form a sample of size  $2N$ . This means that the expected value of the plug-in estimator for a sample of size  $N$  must be less than or equal to the expected value for a sample of size  $2N$ . The conclusion follows.

<sup>12</sup> The larger sample gives a smoother frequency distribution.

<sup>13</sup> The maximum and minimum values for the fitted normal distribution are defined by  $\mu \pm \sigma \cdot \Phi^{-1}(\kappa)$  where  $\kappa = 1/(N + 1)$  and  $N = 10,000$  is the sample size.

Figure 1 (Statistics)

**Monte Carlo Simulation Without Variance Reduction**  
**Analysis of CTE(95%) and VaR(95%) for a 10-year European Put Option,  $X = \$110$ ,  $S = \$100$**

	$\hat{C}TE(95\%)$	Normal	$FSE(\hat{C}TE)$	$VaR(95\%)$	Normal	$FSE(VaR)$
Minimum	7.14	7.57	0.99	0.00	-2.09	0.01
1% Percentile	9.80	9.87	1.23	0.18	0.38	0.48
5% Percentile	10.97	10.99	1.34	1.49	1.58	0.73
Median	13.71	13.70	1.64	4.52	4.49	1.75
Average	13.70	13.70	1.64	4.49	4.49	1.86
95% Percentile	16.40	16.41	1.95	7.42	7.41	3.35
99% Percentile	17.41	17.53	2.09	8.58	8.61	4.23
Maximum	19.59	19.83	2.33	10.92	11.08	6.89
Std Deviation	1.65	1.65	0.19	1.77	1.77	0.82
Skewness	-0.06	0.00	0.13	-0.01	0.00	0.89
Kurtosis	-0.03	0.00	0.00	-0.08	0.00	1.55
Relative Error	0.12			0.39		

able). It is interesting to note that for this example of a *light-tailed* distribution (i.e., the discounted value of the payoff under an “in-the-money” European put option),  $\hat{C}TE(95\%)$  is very nearly normally distributed.

We stated earlier that the *CTE* plug-in estimator is always biased low for any finite sample size. The table that accompanies Figure 1 provides convincing, albeit empirical, evidence of this bias. The observed standard deviation for  $\hat{C}TE(95\%)$  is 1.65 for our sample of size 10,000 ( $n = 1,000$  trials for each sample and hence 50 values are used to estimate  $\hat{C}TE(95\%)$  in each case). Therefore, the standard deviation for the mean  $\hat{C}TE(95\%)$  should be close to  $1.65/\sqrt{10,000} = 0.0165$ . At 13.70, our sample mean is approximately six standard deviations from the true value of 13.80—an unlikely occurrence if there were no bias!

The same model was run again<sup>14</sup> assuming a strike price of 90 rather than 110. With the option so far out of the money the estimated *VaR* turns out to be zero on every trial. Table 2 summarizes the results.

We see that the average *CTE* has dropped by a factor of 3 while the estimated error has only decreased by a factor of 1.6. That is, the *relative* sampling error is now much higher.

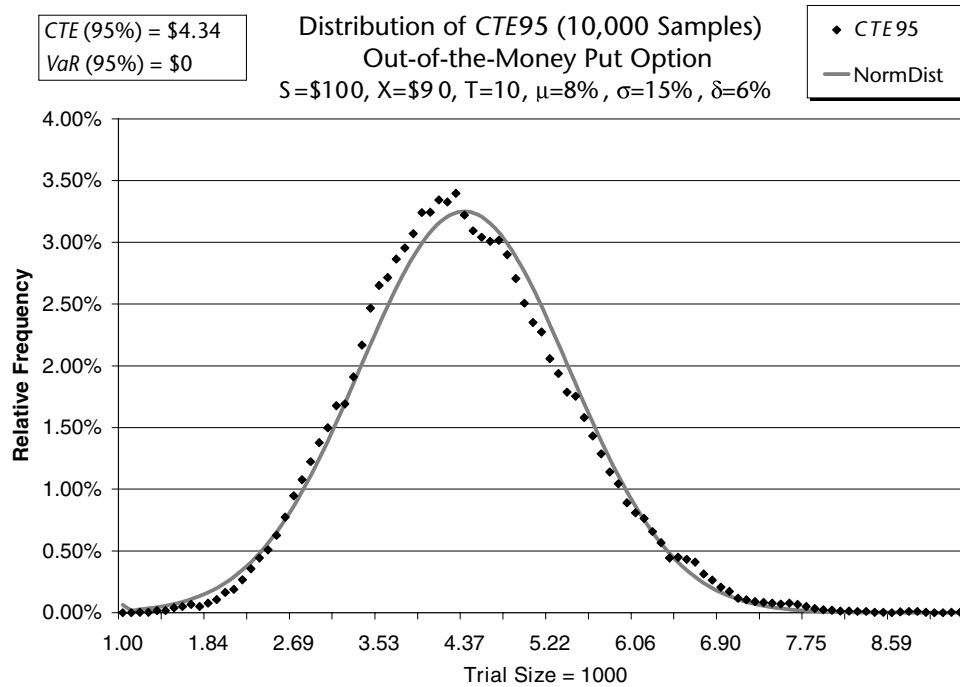
Table 2

**Monte Carlo Simulation Without Variance Reduction**  
**CTE(95%) for a 10-year European Put Option (1000 Trials),  $X = \$90$ ,  $S = \$100$**

	$\hat{C}TE(95\%)$	$FSE(\hat{C}TE)$	$VaR$	$FSE(VaR)$	$Cov(CTE, VaR)$	$\hat{f}(VaR)$
Closed Form	4.34	n/a	0.00	n/a	0.00	n/a
First Trial	4.71	1.13	0.00	n/a	4.53	n/a
Last Trial	3.79	0.96	0.00	n/a	3.60	n/a
Minimum	1.34	0.44	0.00	n/a	1.28	n/a
Average	4.32	1.02	0.00	n/a	4.11	n/a
Maximum	7.58	1.60	0.00	n/a	7.20	n/a
Std Deviation	1.04	0.17	0.00	n/a	1.02	n/a

<sup>14</sup> The random number generator was intentionally “reseeded” when running each new model. For example, the models underlying Tables 1 and 2 use different random numbers.

Figure 2  
**Monte Carlo Simulation Without Variance Reduction**  
**CTE(95%) for a 10-year European Put Option (10,000 Trials),  $X = \$90$ ,  $S = \$100$**



An analysis of the terms contributing to  $FSE(\hat{CTE})$  shows that the first dominates, indicating a heavy-tailed risk. In fact, the first term is usually twice as large as the second. As before, the asymptotic formula for the variance of the CTE estimator closely approximates the empirical variance on average.

Figure 2 provides the empirical distribution of  $\hat{CTE}(95\%)$  for 10,000 samples. For this fatter-tailed risk,  $\hat{CTE}(95\%)$  is only approximately normally distributed. We will return to this example later in the section on variance reduction.

Figure 2 (Statistics)  
**Monte Carlo Simulation Without Variance Reduction**  
**Analysis of CTE(95%) and VaR(95%) for a 10-year European Put Option,  $X = \$90$ ,  $S = \$100$**

	$\hat{CTE}(95\%)$	Normal	$FSE(\hat{CTE})$	$VaR(95\%)$	Normal	$FSE(VaR)$
Minimum	1.22	0.48	0.40	0.00	-0.06	0.01
1% Percentile	2.18	1.92	0.63	0.00	-0.04	0.48
5% Percentile	2.72	2.63	0.75	0.00	-0.03	0.73
Median	4.27	4.33	1.03	0.00	0.00	1.75
Average	4.33	4.33	1.03	0.00	0.00	1.86
95% Percentile	6.12	6.03	1.32	0.00	0.03	3.35
99% Percentile	6.93	6.74	1.43	0.00	0.04	4.23
Maximum	9.25	8.18	1.77	1.44	0.06	6.89
Std Deviation	1.03	1.03	0.17	0.02	0.02	0.82
Skewness	0.31	0.00	0.00	80.24	0.00	0.89
Kurtosis	0.16	0.00	-0.03	6,754.29	0.00	1.55
Relative Error	0.24			74.41		

Figure 3  
**CTE Estimator Versus Quantile (VaR) Estimator at  $\alpha = 0.95$  Level  
 10-year European Put Option (1000 Trials),  $X = \$110$ ,  $S = \$100$**

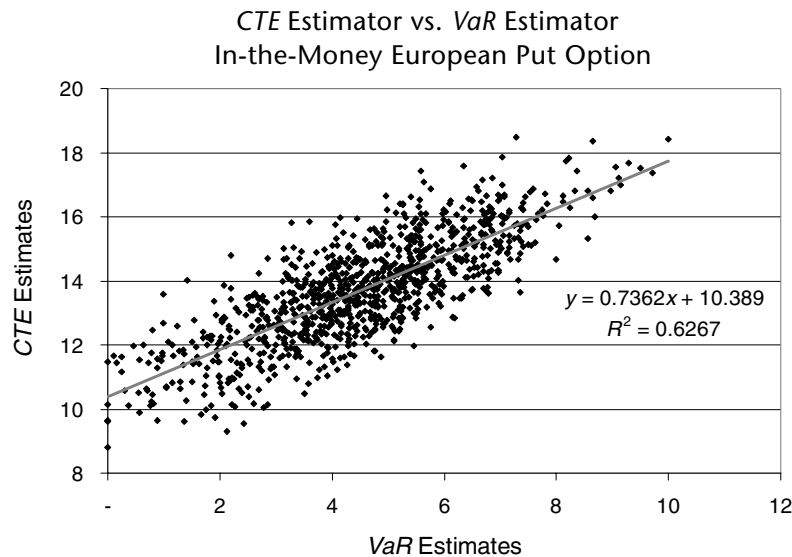


Figure 3 shows a scatter plot for the empirical joint distribution of the *CTE* and *VaR* estimators and the regression of the *CTE* on the *VaR* estimator for the example in Table 1 (i.e., “in-the-money” European put option). The results are consistent with the theory described earlier in this section. Specifically, the two estimators are positively correlated.

### 3. SAMPLING FROM THE PARETO DISTRIBUTION

This section is designed to answer the question of whether the asymptotic formula for the variance of the *CTE* estimator is good enough to work in practice. We do this by seeing how well the formula performs when the variable  $X$  has a Pareto distribution. The Pareto distribution is tractable enough that we can obtain a closed form expression for the variance of the *CTE* estimator for finite sample sizes. Furthermore, the Pareto is applicable to a wide variety of loss distributions and can be a representative proxy for many real-world data problems. As we shall see, the analysis suggests that the asymptotic formula will often work well for reasonable sample sizes.

The distribution function for a Pareto random variable  $X$  is given by

$$F(x) = 1 - \left( \frac{\lambda}{\lambda + x} \right)^\alpha, \quad x > 0.$$

The parameters  $\alpha$  and  $\lambda$  are sometimes called the *shape* and *scale* parameters, respectively. Clearly, the random variable  $Y \equiv X - \psi$  is also Pareto provided  $x > \psi$ , so we may translate the distribution to suit the data. Without loss of generality, we will ignore the location parameter and use  $\psi = 0$  in the following analysis.

It will be convenient to use the transformation  $\xi = 1/\alpha$  and  $\beta = \lambda/\alpha$  so that the distribution function for the Pareto is given by

$$F(x) = 1 - \left( \frac{\beta}{\beta + \xi \cdot x} \right)^{\frac{1}{\xi}}, \quad x > 0.$$

The Pareto distribution is usually referred to as ‘‘heavy-tailed’’ if  $\xi > 0$  and light-tailed otherwise. The mean and variance of  $X|X > u$  are given by

$$E[X|X > u] = \frac{\beta + u}{1 - \xi}, \quad \xi < 1$$

$$VAR[X|X > u] = \frac{(\beta + \xi \cdot u)^2}{(1 - 2\xi) \cdot (1 - \xi)^2}, \quad \xi < 1/2.$$

For a sample of size  $n$ , it then follows that the expected value of the CTE estimator is

$$E[C\hat{T}E_n] = E[E(X|X > x_{(k+1)})],$$

$$= E\left[\frac{\beta + x_{(k+1)}}{1 - \xi}\right],$$

$$= \frac{\beta + E[x_{(k+1)}]}{1 - \xi},$$

and for the variance of the estimator we get

$$VAR(C\hat{T}E_n) = E[VAR(X|X > x_{(k+1)})] + VAR[E(X|X > x_{(k+1)})],$$

$$= E\left[\frac{(\beta + \xi \cdot x_{(k+1)})^2}{(1 - 2\xi) \cdot (1 - \xi)^2}\right] + VAR\left[\frac{\beta + x_{(k+1)}}{1 - \xi}\right],$$

$$= \frac{\beta^2 + 2\beta\xi \cdot E[x_{(k+1)}] + \xi^2 \cdot E[x_{(k+1)}^2]}{(1 - 2\xi) \cdot (1 - \xi)^2} + \frac{VAR[x_{(k+1)}]}{(1 - \xi)^2}.$$

The moments of the order statistics, can be obtained<sup>15</sup> from the formula for the Pareto distribution. The  $m^{th}$  moment of the  $k^{th}$  order statistic is given by

$$E[X_{(k)}^m] = n \cdot \binom{n-1}{k-1} \int_{-\infty}^{\infty} x^m f(x) [F(x)]^{n-k} [1 - F(x)]^{k-1} dx$$

$$= n \cdot \binom{n-1}{k-1} \int_0^1 [F^{-1}(z)]^m z^{n-k} [1 - z]^{k-1} dz$$

$$= \left(\frac{\beta}{\xi}\right)^m \cdot n \cdot \binom{n-1}{k-1} \int_0^1 [(1 - z)^{-\xi} - 1]^m z^{n-k} [1 - z]^{k-1} dz.$$

For a given  $m$  this can be broken down into a tractable expression by using the gamma function in the well-known formula

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 u^{a-1}(1-u)^{b-1} du.$$

For example, if  $m = 1$  we find

<sup>15</sup> See Panjer and Willmot (1992), chapter 4.

$$\begin{aligned}
 E[X_{(k)}] &= \left(\frac{\beta}{\xi}\right) \cdot n \cdot \binom{n-1}{k-1} \int_0^1 [(1-z)^{-\xi} - 1] z^{n-k} [1-z]^{k-1} dz, \\
 &= \left(\frac{\beta}{\xi}\right) \cdot n \cdot \binom{n-1}{k-1} \int_0^1 [z^{n-k} \cdot (1-z)^{k-1-\xi} - z^{n-k}(1-z)^{k-1}] dz, \\
 &= \left(\frac{\beta}{\xi}\right) \cdot n \cdot \binom{n-1}{k-1} \cdot \left[ \frac{\Gamma(n-k+1)\Gamma(k-\xi)}{\Gamma(n+1-\xi)} - \frac{\Gamma(n-k+1)\Gamma(k)}{\Gamma(n+1)} \right], \\
 &= \left(\frac{\beta}{\xi}\right) \cdot \left[ \frac{\Gamma(n+1)\Gamma(k-\xi)}{\Gamma(n+1-\xi)\Gamma(n-k+1)} - 1 \right].
 \end{aligned}$$

A similar, but more complex expression is found for  $m = 2$ :

$$E[X_{(k)}^2] = \left(\frac{\beta}{\xi}\right)^2 \cdot \left( \frac{\Gamma(k-2\xi)\Gamma(n+1)}{\Gamma(k)\Gamma(n+1-\xi)} - 2 \cdot \frac{\Gamma(k-\xi)\Gamma(n+1)}{\Gamma(k)\Gamma(n+1-\xi)} + 1 \right).$$

Thus, we can derive closed form expressions for the mean and variance of the *CTE* estimator in terms of the gamma function. The resulting expressions are cumbersome, but can easily be evaluated in electronic spreadsheets.

In the following examples, we assume that the underlying loss distribution (i.e., risk exposure) is Pareto<sup>16</sup> with  $\beta = 10$  and examine the *CTE* estimator at the  $\alpha = 0.95$  confidence level.<sup>17</sup>

First, we examine the relative bias of the estimator for various sample sizes. From Table 3, we see that there is some negative bias for small sample sizes, but this quickly becomes negligible when  $n \geq 1000$ . Notably, the bias is much smaller at the lower  $\alpha$ -levels. For example, the bias at the  $\alpha = 0.80$  level is about 25–50% of the amounts shown in Table 3.

Table 3

**CTE Estimator Bias from the Pareto Distribution with  $\beta = 10$  and  $\alpha = 0.95$**

$\xi$	Mean	Stdev	<i>CTE</i> ( $\alpha$ )	Estimator Bias by Sample Size			
				100	500	1,000	10,000
-0.4	7.14	5.32	19.61	-1.41%	-0.29%	-0.15%	-0.01%
-0.3	7.69	6.08	22.90	-1.63%	-0.34%	-0.17%	-0.02%
-0.2	8.33	7.04	27.11	-1.85%	-0.38%	-0.19%	-0.02%
-0.1	9.09	8.30	32.62	-2.08%	-0.43%	-0.22%	-0.02%
0.0	10.00	10.00	39.96	-2.29%	-0.47%	-0.24%	-0.02%
0.1	11.11	12.42	49.92	-2.48%	-0.51%	-0.26%	-0.03%
0.2	12.50	16.14	63.79	-2.63%	-0.54%	-0.27%	-0.03%
0.3	14.29	22.59	83.64	-2.72%	-0.56%	-0.28%	-0.03%
0.4	16.67	37.27	113.10	-2.73%	-0.55%	-0.28%	-0.03%

Next, we consider the relative sampling error, defined as the standard error of the *CTE* estimator as a fraction of the estimator itself. That is, for various sample sizes, we look at the quantity

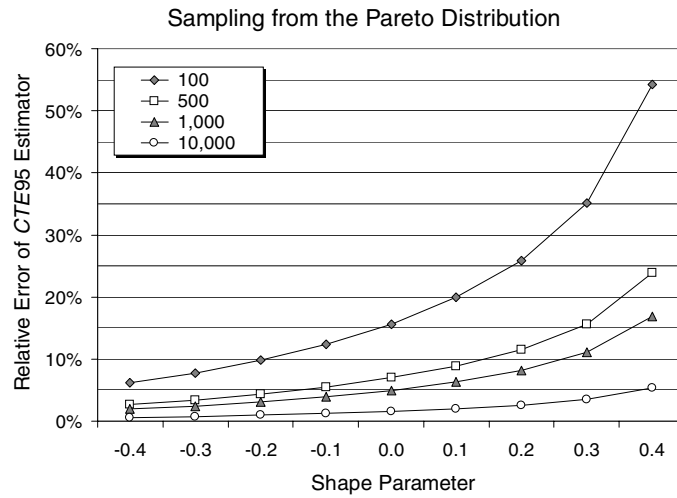
$$\frac{\sqrt{\text{VAR}(\hat{CTE})}}{\hat{CTE}}.$$

From Figure 4, we see that the relative error can be substantial for small samples. Not unexpectedly, the relative error is also higher for the heavy-tailed distributions (i.e., higher values of  $\xi$ ). Even with

<sup>16</sup> In our experience,  $\beta = 10$  is not an unreasonable assumption for many approximate Pareto loss distributions in life insurance, such as benefit claims for investment guarantees on equity-linked variable annuities.

<sup>17</sup> *CTE*(95%) is the prescribed risk measure for the minimum total balance sheet provision (i.e., reserves + capital) in respect of investment guarantee risk on segregated fund variable insurance products in Canada.

Figure 4  
**Relative Error from the Pareto Distribution with  $\beta = 10$  and  $\alpha = 0.95$**



$n = 1,000$  there can be significant sampling error when  $\xi > 0$ . As with bias, the relative error is smaller for lower  $\alpha$ -levels.

It is also instructive to examine the relative contribution of each component in the formula

$$\begin{aligned} \text{VAR}(\hat{CTE}) &= E[\text{VAR}(\hat{CTE}|X_{(k+1)})] + \text{VAR}[E(\hat{CTE}|X_{(k+1)})], \\ &\approx \frac{\text{VAR}(X_{(1)}, \dots, X_{(k)}) + \alpha \cdot (\hat{CTE} - X_{(k)})^2}{k}. \end{aligned}$$

For any distribution with finite variance we can define a tail shape function  $\xi(\alpha)$  by the relation

$$1 - 2 \cdot \xi(\alpha) = \frac{[CTE(\alpha) - VaR(\alpha)]^2}{\text{VAR}[X|X > VaR(\alpha)]}.$$

It can be verified that for the Pareto distribution this quantity is a constant i.e.,  $\xi(\alpha) = \xi$  for all  $0 \leq \alpha \leq 1$ . If we now define a ‘‘Gross Up’’ factor by

$$G_{n,\alpha} = \frac{\text{VAR}[E(\hat{CTE}|X_{(k+1)})]}{E[\text{VAR}(\hat{CTE}|X_{(k+1)})]}.$$

That is,

$$\text{VAR}(\hat{CTE}) = \frac{1}{k} \cdot \text{VAR}[X_{(1)}, \dots, X_{(k)}] \times (1 + G_{n,\alpha}),$$

then we can deduce that for the Pareto distribution, we have the asymptotic result

$$\lim_{n \rightarrow \infty} G_{n,\alpha} = \alpha \cdot (1 - 2\xi).$$

From Table 4, we see that the asymptotic value ( $n \rightarrow \infty$ ) is valid for moderate sample sizes when drawing from a Pareto distribution. Clearly,  $G_{n,\alpha}$  decreases as  $\xi$  increases, meaning that the first term  $\text{VAR}(x_{(1)}, \dots, x_{(k)})$  dominates  $\text{VAR}(\hat{CTE})$  for heavy-tailed distributions. This may provide us some insight when dealing with empirical distributions that are not Pareto, which we shall explore in the next section.

Table 4  
 $G_{n,\alpha}$  for the Pareto Distribution with  $\beta = 10$  and  $\alpha = 0.95$

$\xi$	Sample Size				
	100	500	1,000	10,000	$\infty$
-0.4	1.416	1.642	1.676	1.706	1.710
-0.3	1.288	1.468	1.493	1.517	1.520
-0.2	1.152	1.291	1.310	1.328	1.330
-0.1	1.008	1.111	1.125	1.138	1.140
0.0	0.857	0.931	0.941	0.949	0.950
0.1	0.698	0.747	0.753	0.759	0.760
0.2	0.533	0.562	0.566	0.569	0.570
0.3	0.361	0.376	0.378	0.380	0.380
0.4	0.183	0.189	0.189	0.190	0.190

#### 4. FORMULA RESULTS VERSUS REAL-WORLD SIMULATIONS

The previous section provided some insight into how well the *CTE* variance formula works when the underlying risk follows a Pareto distribution. Of course, the empirical results for most real-world insurance company models would not be exactly (perhaps not even approximately) Pareto, so it is worthwhile to explore how well the formula works for more realistic examples obtained by Monte Carlo simulation.

In all cases, we generate a sample from a real-world problem and use the data to estimate both the *CTE* and its variance using the asymptotic formula. The sampling process is then repeated many times, providing a range of *CTE* estimates. This range is then compared to the results derived from the original samples and the formula. The examples support the use of the formula as a useful measure of sampling error.

We have selected the following test cases for review.

- Guaranteed minimum death and maturity benefits payable on equity-linked variable annuity contracts. The loss distribution is defined by the actuarial present value of benefit claims in excess of the account value. Uncertainty in the loss distribution is modelled by assuming that equity returns follow a regime-switching lognormal process with two states (RSLN2). See Hardy (2001) for a thorough description of the RSLN2 model. All other variables (e.g., mortality, policy lapsation, interest rates, etc.) are known and deterministic.
- A diversified portfolio of whole life insurance with guaranteed premiums and fixed cash values. The loss distribution is defined by the function for gross premium reserves; that is, the actuarial present value of benefits and expenses less policy premiums. Only mortality is stochastic; all other variables (e.g., policy lapse, expenses, interest rates, etc.) are known and deterministic. In each period and for each policy, a random draw from the  $U(0,1)$  distribution is compared to the assumed mortality table to determine if the contract terminates by the death of the insured.
- Aggregate operational risk exposure over one year for multiple independent classes of events. The claims for each class are modelled by a compound distribution with Poisson frequency and lognormal severity.

Our first example, commonly called a “segregated fund variable insurance contract” in Canada, will receive special attention due to its close similarity to the European put option examined in Section 2 and explored again in the next section on variance reduction. The policy is sold as a single premium insurance with a term of 10 years. The deposit is fully invested in a diversified equity fund held at market value. The policy offers a return-of-principal guarantee upon the death of the insured or at the maturity date. That is, at death or maturity, the policyholder will receive the higher of the account value and the original premium (i.e., the insured holds a put option). The contract is typically guaranteed-renewable and “rolls over” into a new policy at the end of 10 years. The company charges the policyholder a fixed spread (% of market value) each month to cover expenses, benefit claims, cost-

Table 5  
**Sample Statistics for Real-World Case Studies**

	At Issue Seg Fund T = 10	In-the-Money Seg Fund T = 6	Whole Life Insurance	Operational Risk
N	60,000	60,000	60,000	60,000
Minimum	0.00	0.00	0.73	0.43
1% Percentile	0.00	0.01	0.83	0.62
5% Percentile	0.00	0.02	0.88	0.70
Median	0.16	0.20	1.00	0.94
95% Percentile	5.34	4.22	1.12	1.42
99% Percentile	9.37	5.63	1.18	2.27
Maximum	27.24	8.30	1.31	15.13
Average	1.00	1.00	1.00	1.00
CTE(70%)	3.04	2.84	1.09	1.31
CTE(95%)	7.78	5.08	1.16	2.03
CTE(99%)	11.02	6.17	1.20	3.48
Std Deviation	1.94	1.42	0.07	0.37
Skewness	3.02	1.67	0.07	9.24
Kurtosis	10.84	2.06	0.01	174.18

of-capital, and profit. Cash flows are determined using realistic assumptions for mortality and policy surrender. Discounting is at a level fixed rate that does not vary by scenario.

Table 5 shows some sample statistics for each loss distribution. There are two examples for the segregated fund policy: “at issue” (the contract guarantee is “at-the-money”) and “in-the-money” (the market value is 75% of the guaranteed value, time-to-maturity = 6 years). For ease of comparison, all the examples have been engineered (scaled) to have an average (mean) of 1. We will focus on CTE(95%), except for the operational risk example where we look at CTE(99%) due to the short (1 year) time horizon.

Before we examine how well the CTE variance formula behaves, it is worth noting that each of these distributions is quite distinct from the others. The two “segregated fund” examples have fat-tails, but the “at issue” (market value = guaranteed value) policy is more skewed and much thicker-tailed. The loss distribution for the mortality risk (whole life insurance) is very nearly normally distributed and the operational risk profile is extremely fat-tailed.

Tables 6 through 9 provide the sample statistics for the simulated loss distributions.

Table 6  
**Equity Risk—Monte Carlo Simulation Without Variance Reduction  
 CTE(95%) for an “At Issue” Segregated Fund (1000 Trials), MV = GV**

	$\hat{CTE}(95\%)$	$FSE(\hat{CTE})$	$VaR$	$FSE(VaR)$	$Cov(CTE, VaR)$	$\hat{f}(VaR)$
Exact	7.78	n/a	5.34	n/a	0.14	n/a
First Trial	7.82	0.48	5.17	0.56	0.20	0.0124
Last Trial	7.33	0.40	5.11	0.12	0.04	0.0558
Minimum	6.26	0.29	4.24	0.05	0.02	0.0057
Average	7.78	0.44	5.38	0.43	0.15	0.0200
Maximum	9.11	0.74	6.84	1.22	0.49	0.1272
Std Deviation	0.44	0.06	0.40	0.19	0.07	0.0118

The real-world examples support a number of practical insights:

- Any given trial gives a reasonable estimate for the CTE and VaR risk measures, although sampling error can be significant. Although we have not shown the results, when  $n < 1000$  and the risk profile is fatter-tailed than the normal distribution, the sampling error can be so large that it renders any tail measures of questionable value for decision making.

Table 7

**Equity Risk—Monte Carlo Simulation Without Variance Reduction  
CTE(95%) for an “In-the-Money” Segregated Fund (1000 Trials), MV = 75% × GV**

	$\hat{CTE}(95\%)$	$FSE(\hat{CTE})$	$VaR$	$FSE(VaR)$	$Cov(CTE, VaR)$	$\hat{f}(VaR)$
Exact	5.08	n/a	4.22	n/a	0.02	n/a
First Trial	4.92	0.16	3.98	0.14	0.02	0.0479
Last Trial	5.03	0.18	3.99	0.33	0.05	0.0209
Minimum	4.50	0.08	3.75	0.01	< 0.01	0.0143
Average	5.07	0.15	4.23	0.16	0.02	0.0523
Maximum	5.54	0.20	4.76	0.48	0.06	0.5833
Std Deviation	0.15	0.02	0.16	0.07	0.01	0.0326

Table 8

**Mortality Risk—Monte Carlo Simulation Without Variance Reduction  
CTE(95%) for a Diversified Whole Life Insurance Portfolio (1000 Trials)**

	$\hat{CTE}(95\%)$	$FSE(\hat{CTE})$	$VaR$	$FSE(VaR)$	$Cov(CTE, VaR)$	$\hat{f}(VaR)$
Exact	1.16	n/a	1.12	n/a	< 0.01	n/a
First Trial	1.16	0.01	1.12	0.01	< 0.01	1.1309
Last Trial	1.15	0.01	1.12	< 0.01	< 0.01	3.1136
Minimum	1.14	0.00	1.11	< 0.01	< 0.01	0.4154
Average	1.15	0.01	1.12	0.01	< 0.01	1.6556
Maximum	1.18	0.01	1.14	0.02	< 0.01	11.2167
Std Deviation	0.01	< 0.01	< 0.01	< 0.01	< 0.01	0.9902

Table 9

**Operational Risk—Monte Carlo Simulation Without Variance Reduction  
CTE(99%) Poisson Frequency, Lognormal Severity (1000 Trials)**

	$\hat{CTE}(95\%)$	$FSE(\hat{CTE})$	$VaR$	$FSE(VaR)$	$Cov(CTE, VaR)$	$\hat{f}(VaR)$
Exact	3.48	n/a	2.27	n/a	0.10	n/a
First Trial	3.38	0.33	2.70	0.32	0.07	0.0097
Last Trial	2.83	0.36	2.09	0.23	0.05	0.0134
Minimum	2.13	0.08	1.72	0.04	0.01	0.0015
Average	3.41	0.54	2.33	0.47	0.19	0.0092
Maximum	5.83	1.62	3.42	2.08	1.72	0.0894
Std Deviation	0.59	0.26	0.28	0.27	0.18	0.0062

- Bias, if any, in the  $CTE$  or  $VaR$  estimators is small relative to the sampling error.
- Statistics for “nearly normal” distributions can be estimated with high precision with relatively small samples.
- The asymptotic  $CTE$  variance formula is an excellent approximation for the true (empirical) variance, even for high  $\alpha$ -levels. In all of the examples,  $E[FSE(\hat{CTE})] \approx \sigma(\hat{CTE})$ .
- On average, the standard error formula for the quantile (i.e.,  $VaR$ ) estimator overstates the true (empirical) standard deviation when the loss distribution is heavy-tailed. That is,  $E[FSE(VaR)] > \sigma(VaR)$ . Furthermore,  $FSE(VaR)$  can significantly misestimate the true standard deviation. This error is primarily due to difficulties in estimating  $f(VaR)$  from the empirical distribution.

## 5. VARIANCE REDUCTION TECHNIQUES

This section discusses the important subject of sampling error. If an application of the formula shows that the sampling error is “too large,” then a practitioner has only a few alternatives. One obvious

choice is to increase the sample size, but this could prove impractical since the requisite number of samples to enhance accuracy may be prohibitively large. A more efficient approach involves the use of a “control portfolio,” an intuitive and flexible alternative that is closely aligned with the control variate method<sup>18</sup> of variance reduction. An example of one such approach is presented at the end of Section 6.

Another more robust approach to variance reduction uses a system of biased sampling to improve the quality of the CTE estimate by putting more sample elements in the tail of the distribution that is being measured. We show how the formulas for estimating variance need to be adjusted to account for bias in the sampling method.

One of the most important variance reduction techniques in Monte Carlo simulation is an approach known as *importance sampling*. In this approach, the actual sampling distribution  $F$  is replaced by a new sampling distribution  $G$  such that the density functions  $dF$ ,  $dG$  are related by an everywhere positive weight function  $W$  so that  $dF = WdG$ . A random sample of size  $n$  is then drawn from the distribution  $G$  and estimators for the CTE and VaR are determined from

$$1 - \alpha = \int_{VaR_n}^{\infty} Wd\hat{G}_n,$$

$$CTE\hat{E}_n = \frac{1}{1 - \alpha} \int_{VaR_n}^{\infty} x \cdot Wd\hat{G}_n.$$

The sampling method has to be engineered so that the weight function  $W = W(x)$  is a *known function of the sample*. An example will be given later.

The delta method described earlier can be extended to this situation. The expressions for the influence functions in this more general setting are

$$IF_{VaR}(x) = \begin{cases} \frac{-(1 - \alpha)}{f(VaR)} & x < VaR \\ \frac{W(x) - (1 - \alpha)}{f(VaR)} & x > VaR \end{cases}$$

and

$$IF_{CTE}(x) = \begin{cases} VaR - CTE & x < VaR \\ VaR - CTE + \frac{W(x)[x - VaR]}{1 - \alpha} & x > VaR \end{cases}.$$

Both of these expressions can be simplified by using the notation

$$H(x) = \begin{cases} 0 & x < VaR, \\ 1 & x \geq VaR. \end{cases}$$

So that

$$IF_{VaR}(x) = \frac{W(x)H(x) - (1 - \alpha)}{f(VaR)},$$

$$IF_{CTE}(x) = \frac{W(x)H(x)(x - VaR) + (1 - \alpha)(VaR - CTE)}{(1 - \alpha)}.$$

The delta method then gives us the following large sample approximations for the variances of the plug-in estimators

<sup>18</sup> See Hardy (2003) for a brief, but eminently understandable discussion.

$$\begin{aligned} \text{VAR}(\hat{VaR}_n) &\approx \frac{E_G[(IF_{VaR})^2]}{n} = \frac{\text{VAR}_G[WH]}{n \cdot [f(VaR)]^2}, \\ \text{VAR}(\hat{CTE}_n) &\approx \frac{E_G[(IF_{CTE})^2]}{n} = \frac{\text{VAR}_G[W(X - VaR)H]}{n \cdot (1 - \alpha)^2}, \\ \text{Cov}(\hat{CTE}_n, \hat{VaR}_n) &\approx \frac{E_G[IF_{CTE} \cdot IF_{VaR}]}{n} = \frac{\text{Cov}_G[WH, WH(X - VaR)]}{n \cdot f(VaR) \cdot (1 - \alpha)}. \end{aligned}$$

These expressions are convenient for implementing the theory, but it is not easy to see how the importance of the sampling method is affecting the variance. To get a better understanding of the issues we rewrite the above formulas in terms of the true probability distribution  $F$  by using the fact that for any random variable  $X$  we have  $E_F(X) = E_G(WX)$ .

In terms of  $F$  the new expressions are

$$\begin{aligned} \text{VAR}(\hat{VaR}_n) &\approx (1 - \alpha) \cdot \frac{E_F[W|X \geq VaR] - 1 + \alpha}{n \cdot [f(VaR)]^2}, \\ \text{VAR}(\hat{CTE}_n) &\approx \frac{\text{VAR}_F[\sqrt{WH}(X - VaR)]}{n \cdot (1 - \alpha)^2}, \\ &= \frac{E_F[W|X \geq VaR] \cdot \text{VAR}_F[X|X \geq VaR]}{n \cdot (1 - \alpha)} \\ &\quad + \frac{(CTE - VaR)^2 \cdot (E_F[W|X \geq VaR] - 1 + \alpha)}{n \cdot (1 - \alpha)} \\ &\quad + \frac{\text{Cov}_F[W, (X - VaR)^2|X \geq VaR]}{n \cdot (1 - \alpha)} \end{aligned}$$

and

$$\text{Cov}(\hat{CTE}_n, \hat{VaR}_n) \approx \frac{\text{Cov}_F[W, X - VaR|X \geq VaR]}{n \cdot f(VaR)} + \frac{(1 - \alpha)}{n \cdot f(VaR)} \cdot \{E_F[W|X \geq VaR] - 1 + \alpha\}.$$

We recover the earlier versions of these formulas when  $W = 1$ .

The first of these results suggests that the way to reduce the variance of the  $VaR$  estimator is to choose the weighting system so that the quantity  $\{E_F[W|X \geq VaR] - 1 + \alpha\}$  is as small as possible. It is not hard to show this quantity must be positive because

$$\begin{aligned} 0 &\leq \text{VAR}_G[WH], \\ &= E_G[W^2H] - (1 - \alpha)^2, \\ &= (1 - \alpha) \cdot \{E_F[W|X \geq VaR] - (1 - \alpha)\}. \end{aligned}$$

Intuitively, this is reasonable since it is telling us to put more samples in the critical tail of the distribution.

The formula for the  $CTE$  variance suggests that it is also a good idea to engineer the weight function so that it is negatively correlated with the quantity  $(X - VaR)^2$  when  $X > VaR$ .

The final formula for the covariance shows the two estimators become less correlated when a variance reduction scheme is implemented.

We will not try to find a “best” sampling strategy in this paper. We will conclude this section by applying the ideas to the example of a simple European put option used earlier. The example will show that the ideas work in the sense that importance sampling will reduce sampling error and will often reduce estimator bias as well. However, a poor choice of sampling distribution runs the risk of making the situation worse.

For the example of the simple put liability  $Y = e^{-\delta T} \cdot \max[0, X - S \cdot e^{(\mu T + \sigma \sqrt{T} z)}]$  consider the following modified sampling strategy.

1. Generate a random  $N(0,1)$  sample  $z$ .
2. If  $z < \Phi^{-1}(\beta)$  for some  $0 < \beta < 1$  (i.e., the sample is in the tail), we keep the sample.
3. If  $z > \Phi^{-1}(\beta)$  the sample is retained with some small fixed acceptance probability  $p < 1$ .
4. If the above sample is not retained it is mapped to a new sample in the range  $z < \Phi^{-1}(\beta)$  by the transformation

$$z \rightarrow z' = \Phi^{-1} \left[ \beta \cdot \frac{1 - \Phi(z)}{1 - \beta} \right].$$

It can be verified that this mapping corresponds to the weight function

$$W(z) = \begin{cases} \frac{\beta}{1 - p \cdot (1 - \beta)} & z < \Phi^{-1}(\beta), \\ \frac{1}{p} & z \geq \Phi^{-1}(\beta), \end{cases}$$

which is the ratio of the true probability  $dF$  to the sampling probability density  $dG$ .

In addition to importance sampling, statisticians often talk about *stratification* as a variance reduction tool. In the method described above we would expect to get  $n \cdot p \cdot (1 - \beta)$  samples in the range  $z \geq \Phi^{-1}(\beta)$ , but on any given run we may get more or less than that number. We can add stratification to the method above by engineering the process so that we obtain *precisely*  $n \cdot p \cdot (1 - \beta)$  samples from the range  $z \geq \Phi^{-1}(\beta)$  and  $n \cdot [1 - p \cdot (1 - \beta)]$  samples from the range  $z < \Phi^{-1}(\beta)$ . As the examples that follow show, stratification can be important.

With this new methodology, a random sample of size  $n$  is drawn from  $G$  (with or without stratification) and for each sample we have both a value  $X_i$  and a weight  $W_i$ . The  $VaR$  ( $\alpha$ -quantile) is estimated as the highest order statistic  $X_{(i)}$  such that<sup>19</sup>  $\sum_1^i W_j \leq 1 - \alpha$ . We then estimate the  $CTE$  by

$$CTE = \frac{\sum_1^i W_j X_{(j)}}{\sum_1^i W_j}$$

Formula standard errors (FSE) for the estimators are calculated using the formulas developed earlier in this section.

Table 10  
**Importance Sampling  $\beta = 0.06$ ,  $p = 0.10$  Without Stratification**  
**CTE(95%) for a 10-year European Put Option (1000 Trials),  $X = \$110$ ,  $S = \$100$**

	$\hat{CTE}(95\%)$	$FSE(\hat{CTE})$	$VaR$	$FSE(VaR)$	$Cov(CTE, VaR)$	$\hat{f}(VaR)$
Closed Form	13.80	n/a	4.39	n/a	0.04	n/a
First Trial	13.99	0.32	4.42	0.35	0.06	0.0025
Last Trial	13.50	0.30	4.43	0.12	0.02	0.0070
Minimum	12.83	0.29	3.64	0.02	0.00	0.0014
Average	13.79	0.32	4.40	0.23	0.04	0.0041
Maximum	14.86	0.34	5.17	0.63	0.11	1.0070
Std Deviation	0.31	0.01	0.24	0.10	0.02	0.0030

<sup>19</sup> Recall, we have defined the order statistics as  $x_{(1)} \geq x_{(2)} \geq \dots \geq x_{(n)}$ .

Figure 5  
**Importance Sampling  $\beta = 0.06$ ,  $p = 0.10$  Without Stratification**  
**CTE(95%) for a 10-year European Put Option (10,000 Trials),  $X = \$110$ ,  $S = \$100$**

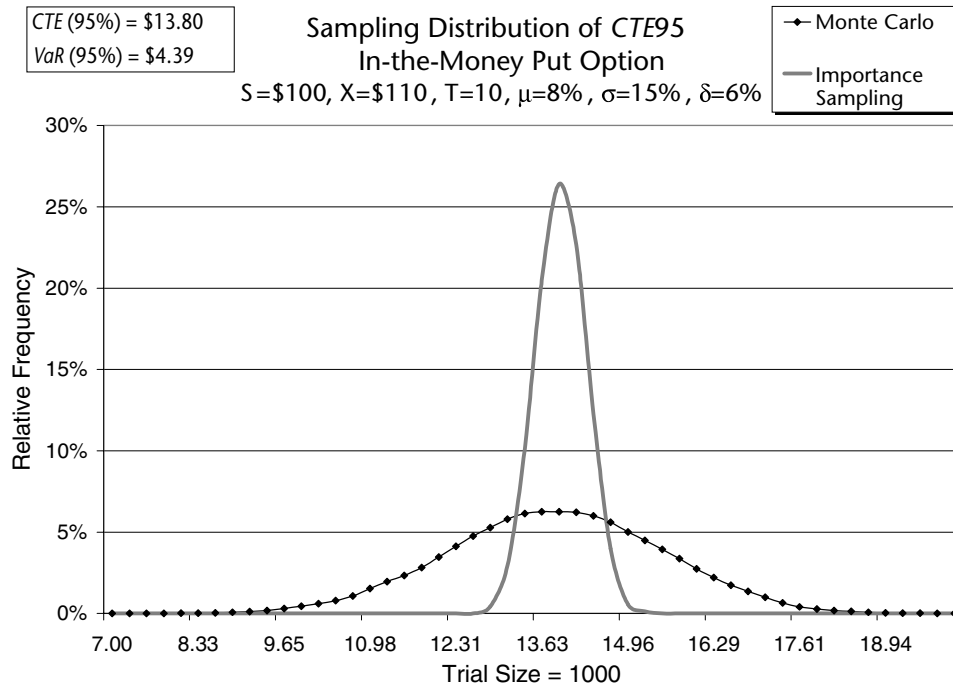


Figure 5 (Statistics)  
**Importance Sampling  $\beta = 0.06$ ,  $p = 0.10$  Without Stratification**

	Monte Carlo Simulation			Importance Sampling Without Stratification		
	$\hat{CTE}(95\%)$	Normal	$FSE(\hat{CTE})$	$\hat{CTE}(95\%)$	Normal	$FSE(\hat{CTE})$
Minimum	7.14	7.57	0.99	12.90	12.63	0.29
1% Percentile	9.80	9.87	1.23	13.06	13.06	0.30
5% Percentile	10.97	10.99	1.34	13.29	13.28	0.30
Median	13.71	13.70	1.64	13.79	13.79	0.32
Average	13.70	13.70	1.64	13.79	13.79	0.32
95% Percentile	16.40	16.41	1.95	14.30	14.31	0.33
99% Percentile	17.41	17.53	2.09	14.51	14.52	0.34
Maximum	19.59	19.83	2.33	14.83	14.96	0.34
Std Deviation	1.65	1.65	0.19	0.31	0.31	0.01
Skewness	-0.06	0.00	0.13	0.03	0.00	0.17
Kurtosis	-0.03	0.00	0.00	-0.10	0.00	0.23
Relative Error	0.12			0.02		

Table 11  
**Importance Sampling  $\beta = 0.05, p = 0.10$  Without Stratification**  
**CTE(95%) for a 10-year European Put Option (1000 Trials),  $X = \$110, S = \$100$**

	$\hat{CTE}(95\%)$	$FSE(\hat{CTE})$	$VaR$	$FSE(VaR)$	$Cov(CTE, VaR)$	$\hat{f}(VaR)$
Closed Form	13.80	n/a	4.39	n/a	0.00	n/a
First Trial	13.87	0.26	4.68	0.11	0.01	0.0048
Last Trial	14.19	0.26	4.40	0.14	0.01	0.0037
Minimum	12.98	0.24	4.09	0.02	0.00	0.0009
Average	13.86	0.26	4.47	0.14	0.01	0.0043
Maximum	14.80	0.28	4.85	0.56	0.06	1.0050
Std Deviation	0.26	0.01	0.12	0.06	0.01	0.0022

Table 10 reports the results for the same “in-the-money” European put option ( $X = \$110, S = \$100$ ) as before (see Table 1), but with weight function parameters  $\beta = 0.06$  and  $p = 0.10$ . Stratification is not used.

The strategy has definitely improved the quality of the estimators for both  $CTE$  and  $VaR$  and the correlation between them has been reduced. To achieve this level of precision without importance sampling (i.e., using straight Monte Carlo without variance reduction) would have required a sample size of about 25,000. The variance approximation formulas are performing well; that is,  $FSE(\hat{CTE}) \approx$  empirical standard deviation of  $\hat{CTE}$ .

Figure 5 provides the frequency distribution of  $\hat{CTE}(95\%)$  using 10,000 trials for the example in Table 10. For reference, the prior results using Monte Carlo simulation (see Figure 1) are also shown. This visual representation makes it abundantly clear just how well the biased sampling has worked—the distribution is much more tightly centered about the mean. As before,  $\hat{CTE}(95\%)$  appears to be almost normally distributed, but with much lower variance.

Table 11 shows what happens when we adopt a slightly more aggressive importance sampling strategy by using the parameters  $\beta = 0.05, p = 0.10$  and still no stratification.

The sampling error has improved, but at the price of introducing bias. The bias is still much smaller than the sampling error, but it seems to be a high price to pay for a very modest reduction in sampling error. The formula approximations continue to perform very well.

Next, we look at the “out-of-the-money” European put option using the less aggressive strategy  $\beta = 0.06, p = 0.10$  and no stratification. The results are in Table 12 (compare to Table 2).

The importance sampling has not been as effective for this “heavy-tailed” case as it was when the option was “in-the-money.” The precision is comparable to a random sample size of about 7,500. There is also some evidence of small sample bias. However, the bias is still much less than the sampling error. We would conclude that importance sampling is still worth the effort in this case, but it is less effective than when the option is “in-the-money.”

Table 12  
**Importance Sampling  $\beta = 0.06, p = 0.10$  Without Stratification**  
**CTE(95%) for a 10-year European Put Option (1000 Trials),  $X = \$90, S = \$100$**

	$\hat{CTE}(95\%)$	$FSE(\hat{CTE})$	$VaR$	$FSE(VaR)$	$Cov(CTE, VaR)$	$\hat{f}(VaR)$
Closed Form	4.34	n/a	0.00	n/a	0.00	n/a
First Trial	4.36	0.20	0.00	0.05	0.00	n/a
Last Trial	4.58	0.24	0.00	0.02	0.00	n/a
Minimum	3.80	0.20	0.00	0.00	0.00	n/a
Average	4.52	0.22	0.00	0.02	0.00	n/a
Maximum	6.10	0.25	0.00	0.12	0.13	n/a
Std Deviation	0.36	0.01	0.00	0.04	0.00	n/a

Table 13  
**Importance Sampling  $\beta = 0.05$ ,  $p = 0.10$  with Stratification**  
 **$CTE(95\%)$  for a 10-year European Put Option (1000 Trials),  $X = \$90$ ,  $S = \$100$**

	$C\hat{T}E(95\%)$	$FSE(C\hat{T}E)$	$V\hat{a}R$	$FSE(V\hat{a}R)$
Closed Form	4.34	n/a	0.00	n/a
First Trial	4.52	0.20	0.00	0.14
Last Trial	4.25	0.19	0.00	0.10
Minimum	3.84	0.18	0.00	0.01
Average	4.38	0.20	0.00	0.07
Maximum	4.99	0.22	0.00	0.15
Std Deviation	0.18	0.01	0.00	0.04

The formula approximation to the estimator variance is not performing as well since  $FSE(C\hat{T}E)$  materially underestimates the “true” standard deviation of  $C\hat{T}E$  (i.e.,  $0.22 \ll 0.36$ ). There appears to be a source of variance in the process that is not captured by the asymptotic formula. Introducing a stratified sampling method can eliminate this discrepancy. Table 13 shows the result of using a stratified sampling method on the “out-of-the-money” European put option using the parameters  $\beta = 0.05$ ,  $p = 0.10$ .

The stratification has eliminated some of the bias and the asymptotic estimate for the variance of the  $CTE$  estimator now appears to work. We can understand these results by looking at the formula for the variance of the  $CTE$  estimator.

Figure 6  
**Importance Sampling  $\beta = 0.05$ ,  $p = 0.10$  with Stratification**  
 **$CTE(95\%)$  for a 10-year European Put Option (10,000 Trials),  $X = \$90$ ,  $S = \$100$**

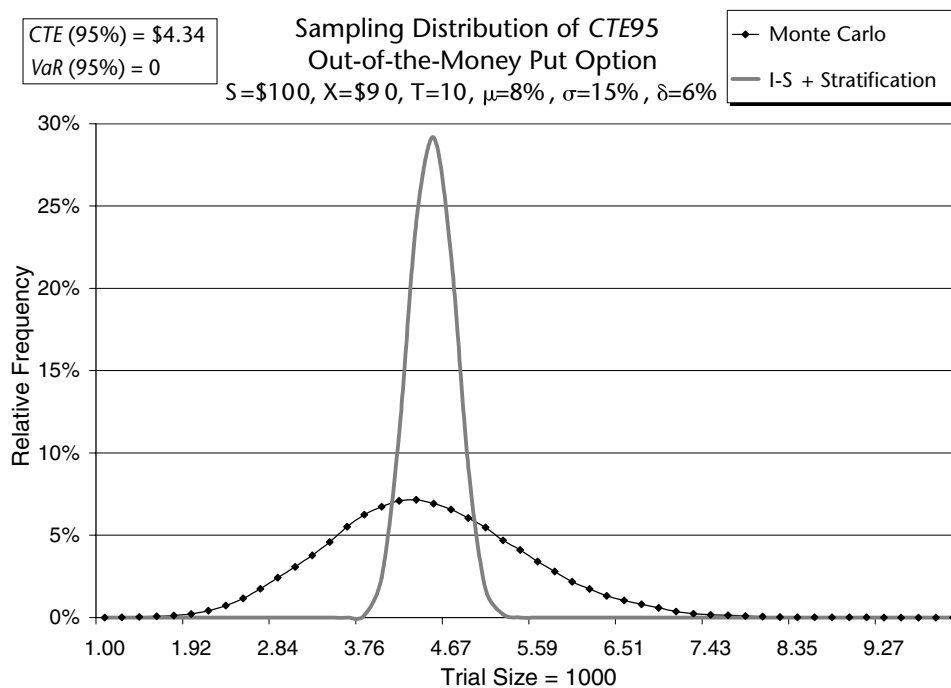


Figure 6 (Statistics)  
**Importance Sampling  $\beta = 0.05$ ,  $p = 0.10$  with Stratification**

	Monte Carlo Simulation			Importance Sampling with Stratification		
	$\hat{CTE}(95\%)$	Normal	$FSE(\hat{CTE})$	$\hat{CTE}(95\%)$	Normal	$FSE(\hat{CTE})$
Minimum	1.22	0.48	0.40	3.86	3.71	0.18
1% Percentile	2.18	1.92	0.63	3.99	3.96	0.18
5% Percentile	2.72	2.63	0.75	4.09	4.09	0.19
Median	4.27	4.33	1.03	4.38	4.38	0.20
Average	4.33	4.33	1.03	4.38	4.38	0.20
95% Percentile	6.12	6.03	1.32	4.68	4.68	0.21
99% Percentile	6.93	6.74	1.43	4.83	4.80	0.22
Maximum	9.25	8.18	1.77	4.97	5.05	0.22
Std Deviation	1.03	1.03	0.17	0.18	0.18	0.01
Skewness	0.31	0.00	0.00	0.14	0.00	0.07
Kurtosis	0.16	0.00	-0.03	0.00	0.00	-0.22
Relative Error	0.24			0.04		

$$\begin{aligned}
 VAR(\hat{CTE}_n) &= \frac{Cov_F[W, (X - VaR)^2 | X \geq VaR]}{n \cdot (1 - \alpha)} \\
 &+ \frac{E_F[W | X \geq VaR] \cdot VAR_F[X | X \geq VaR]}{n \cdot (1 - \alpha)} \\
 &+ \frac{(CTE - VaR)^2 \cdot (E_F[W | X \geq VaR] - 1 + \alpha)}{n \cdot (1 - \alpha)}.
 \end{aligned}$$

When the distribution is light-tailed (i.e.,  $VAR[X | X > VaR] \ll (CTE - VaR)^2$ ), the importance sampling can have a material impact on both the second and third terms above without introducing significant covariance in the first term. If the risk is heavy-tailed, the second term dominates and we see variance reduction by a factor of about  $E[W | X > VaR]$ .

In the strategy being used

$$E[W | X > VaR] = \frac{\beta}{1 - p \cdot (1 - \beta)} = 0.05525,$$

so we expect the standard deviation of the heavy-tailed distribution to be reduced by about  $\sqrt{.05525} = 0.235$  and a larger reduction for the light-tailed risk. This is exactly what happened. For example, comparing the results from Table 13 to Table 2, we see the empirical standard deviation of  $\hat{CTE}$  drop to 0.18 from 1.04, a ratio of 0.173, which is close to 0.235.

The Table 13 results are graphically depicted in Figure 6 as a frequency distribution for  $\hat{CTE}(95\%)$  using 10,000 trials. For comparison, the Monte Carlo results (see Figure 2) are also shown. Again, we have dramatic evidence for the effectiveness of the biased sampling technique and  $\hat{CTE}(95\%)$  still appears to be normally distributed, but with much lower variance.

The main conclusion to draw from this discussion is that importance sampling gives results for a heavy-tailed risk with a precision comparable to a sample size of about

$$\frac{n}{E[W | X > VaR]},$$

but it can do much better for a light-tailed distribution. A second conclusion is that stratification plays a more important role when dealing with heavy-tailed distributions.

## 6. PRACTICAL APPLICATIONS

In this paper we have presented theory and empirical results supporting the use of the asymptotic formula

$$\text{VAR}[\hat{CTE}] \approx \frac{\text{VAR}[X_{(1)}, \dots, X_{(k)}] + \alpha \cdot (CTE - X_{(k)})^2}{k}$$

for the variance of the *CTE* estimator as a good approximation to the true variance. That is,  $E[\hat{VAR}(CTE)] = \sigma^2(CTE)$ . In many ways, the *CTE* is a superior risk measure to the more traditional *Value-at-Risk* (i.e.,  $\alpha$ -quantile), but it is still subject to considerable sampling error when estimated from small samples. The variance formula can assist the practitioner in understanding the potential misestimation for a given sample size, but it also suggests how variance reduction can dramatically improve the quality of the estimators. We have given the variance formula for the general case using a system of biased weights, and demonstrated the effectiveness of importance sampling.

Many authors have written on the qualities of the *CTE* as a risk measure. Our intention here is not to debate the relative merits of the *CTE*, but to recognize its growing acceptance in the life insurance industry<sup>20</sup> as a measure for determining balance sheet provisions (reserves and minimum required capital) whenever the company's risk exposures (i.e., liabilities) are estimated using stochastic methods. Because the *CTE* considers all losses beyond a given  $\alpha$ -quantile, it is particularly well suited for setting "real world" balance sheet liabilities when the risks are unhedged.<sup>21</sup>

However, most insurance liabilities are complex and particularly influenced by the uncertain future behavior of policyholders.<sup>22</sup> As such, closed-form solutions are generally unavailable, and the loss distribution must be estimated using the flexible, but sometimes computationally intensive technique of Monte Carlo simulation. Data storage requirements and processing times quickly become very practical constraints, limiting the number of trials (scenarios) that can be executed.

The ensuing sampling error is not merely of academic interest. Material over- or understatement can lead to undesirable and/or misleading income volatility and possibly ill-informed management decisions. Understanding the potential error in the *CTE* estimates is important if a company is to put balance sheet changes (e.g., from period to period) into perspective. For example, the sampling error in the liabilities could be far less, or more, than (say) the impact of different assumptions for policyholder behavior or the effects of parameter uncertainty and model risk.

It is good practice to give a confidence interval, or at least indicate the standard error, for any statistical estimates obtained by observation or simulation. Without proper communication, point estimates can give a false sense of precision. Understanding sampling error can also be used to advantage, by allowing the company to define a rigorous process for bringing some stability to the balance sheet. For example, at each valuation date the company could define the nominal liabilities at (say)  $CTE(70\%)$ , but the actual balance sheet provisions would be set in the interval  $\hat{CTE}(70\%) \pm FSE(\hat{CTE})$ . This would permit some measure of interperiod "smoothing" based on objective and sensible criteria.<sup>23</sup>

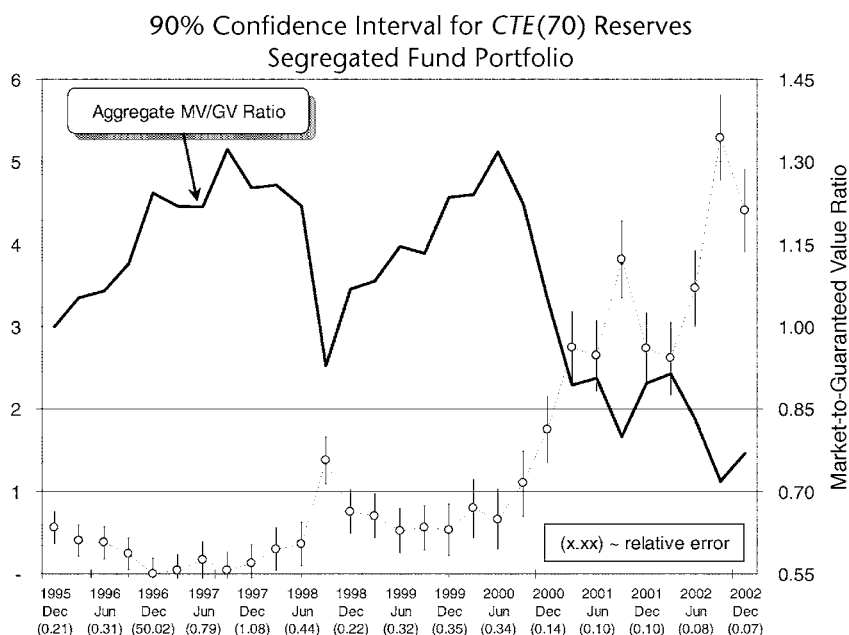
<sup>20</sup> Especially in Canada where actuarial Standards of Practice promulgate the use of the *CTE* whenever stochastic methods are used to set balance sheet liabilities.

<sup>21</sup> It is arguable that when the risk exposure is hedged, the best liability measure is the risk neutral *option value*. An additional provision for parameter misestimation, model error, basis, execution, and liquidity risks (etc.) would be held as required capital. We will not explore this idea further.

<sup>22</sup> We use the term policyholder to represent any customer or entity whose actions cannot be known with a high degree of certainty and whose exercise of contractual privileges has a material impact on the insurer's liabilities.

<sup>23</sup> These criteria could be related to the "tail thickness" of the distribution, but in any event would have to be well defined to avoid manipulation. Otherwise, the company could be induced to always hold the smallest liabilities.

Figure 7  
**Confidence Interval for Nominal CTE(70%) Liabilities**



To explore this idea further, suppose the actual balance sheet liabilities for investment guarantees on variable insurance or annuity contracts are established in the range  $\hat{CTE}(70\% + \Delta)$  with  $-10\% \leq \Delta \leq 10\%$  defined as an objective function of uncertainty<sup>24</sup> related to:

- the valuation data (e.g., “in-the-moneyness” measure); and/or
- the loss distribution (e.g., variance of the CTE(70%) estimator).

To illustrate the concepts,<sup>25</sup> we will track a hypothetical portfolio of 10-year segregated fund policies<sup>26</sup> (i.e., variable annuities) fully invested in Canadian equities over the period December 1995 to December 2002 inclusive. Figure 7 shows an approximate 90% confidence interval for the *nominal* CTE(70%) liabilities, hereafter denoted as  $V_t$ . The aggregate ratio of market-to-guaranteed value (MV/GV) is depicted as a solid line (scale is on the right Y-axis). The relative error, defined as  $\sqrt{\hat{V}AR[\hat{CTE}(70\%)]}/\hat{CTE}(70\%)$ , is shown in parentheses on the X-axis labels. Consistent with prior results, the relative error is higher when the loss distribution is thick-tailed (i.e., the guarantees are out-of-the-money), as occurred up until August 2000 (MV/GV is 1.318). By the end of 2002, the guarantees were deeply in-the-money due to poor equity returns (MV/GV is 0.769).

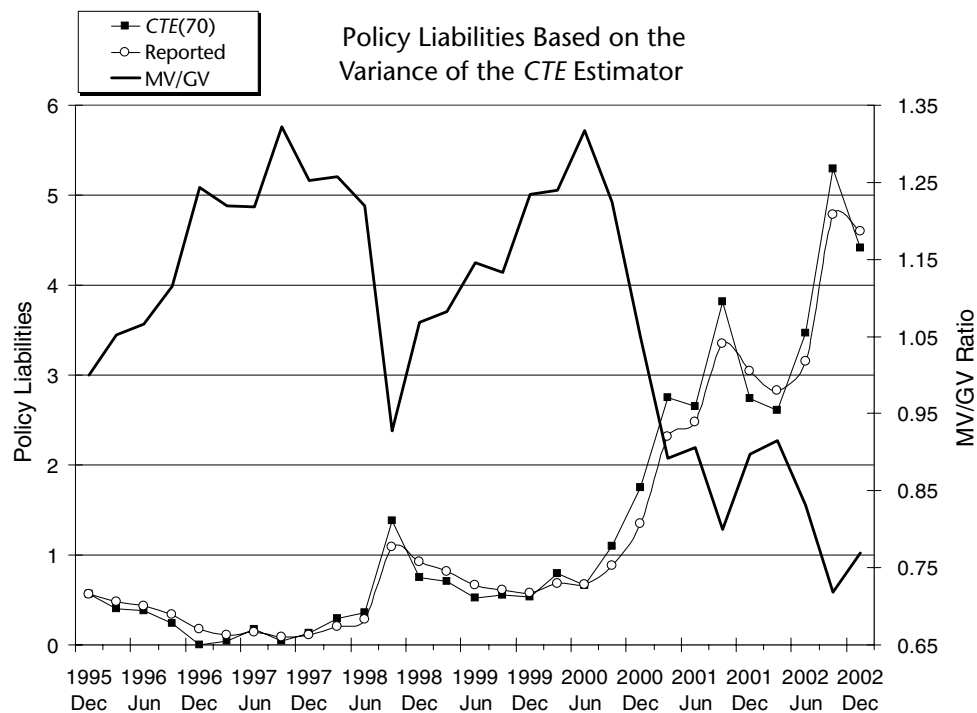
Suppose we now define the *reported* liabilities  $V'_t$  by the formula  $V'_t = V'_{t-1} + 0.5 \times (V_t - V'_{t-1})$  subject to the lower and upper bounds established by the 90% confidence interval for  $V_t$ . The reported

<sup>24</sup> The concept of uncertainty plays an important role in the valuation of policy liabilities under Canadian Generally Accepted Accounting Principles (CGAAP). Specifically, the balance sheet liabilities are “best estimate”, but include a provision for adverse deviation (PfAD) in respect of each of the underlying risk factors. The PfADs are not solvency margins, but rather are related to uncertainty (in the risk factors or the associated loss distribution), with higher provisions required for situations where the actuary is less certain of the potential future financial outcome.

<sup>25</sup> This example follows directly from Hancock (2003).

<sup>26</sup> The product offers an elective reset option (sometimes called a “shout” option in the financial literature) whereby the policyholder can lock-in market gains by “resetting” the guaranteed benefit to the current account value. The maturity date is extended to 10 years after the date of reset (ignoring surrender charges, a “reset” is similar to a lapse and repurchase). To simplify the analysis, we assume a “stationary” attained age 55 population with a constant 5% deposit growth rate.

Figure 8  
**Reported CTE(70%) Liabilities**



liabilities for our hypothetical portfolio are shown in Figure 8. As can be seen, the two “liability paths” are very similar, although the reported values appear less volatile. We can formalize this impression by calculating a measure of “smoothness” for each path. Here, we assume that a cubic polynomial is “smooth” and calculate a *smoothness coefficient* based on traditional 4<sup>th</sup> differences:  $S = \sum_i (\Delta^4 v_i)^2$ . In itself, the measure  $S$  has no particular meaning,<sup>27</sup> but it does allow us to compare the relative smoothness of each path. In this case, the nominal and reported liabilities have smoothness coefficients of 176.6 and 64.2 respectively. The essential point of this analysis is that on the basis of statistical uncertainty, the *reported* liabilities are as reasonable as the *nominal* values, but promote greater stability. Given the uncertainty in estimating liabilities from stochastic simulation, it seems logical not to introduce potentially confusing volatility into the process.

In Section 5, we presented some compelling evidence for the use of simple variance reduction techniques to improve the quality of tail estimators. However, it may not be straightforward or even clear how to implement importance sampling for large-scale Monte Carlo simulations of complex liabilities. Fortunately, there is a very intuitive approach that can yield significant benefits for a given sample size, although the improvement will not be as dramatic as true variance reduction.

While the entire loss (alternatively, profit) profile is often of great interest, certain activities (e.g., the valuation of policy liabilities) naturally focus on the tail of the distribution. For example, suppose a company sets its policy liabilities at  $CTE(70\%)$  and the total balance sheet provision (i.e., reserves plus minimum required capital) at  $CTE(95\%)$ .<sup>28</sup> Running an inforce portfolio over 1000 Monte Carlo scenarios could take many hours and it is certainly wasteful to spend 70% of the processing time on scenarios whose results will be “discarded.”

<sup>27</sup> A lower value of  $S$  denotes a smoother series.

<sup>28</sup> This would be typical of a Canadian life insurance company setting balance sheet provisions for investment guarantee risks on segregated fund variable insurance contracts.

If the company wants to reduce the sampling error in the  $\hat{CTE}$  values for a given number of scenarios, it would be well advised to adopt an approach that shares some similarity to the control variate method of variance reduction.

To conclude this paper, we describe one such solution below.

1. Construct a representative, but suitably small,<sup>29</sup> portfolio that approximates the characteristics of the actual inforce business to be valued. Call this the “control” portfolio.
2. Value the control portfolio over a large<sup>30</sup> number of scenarios  $N$  to give the desired  $CTE$  estimates within an acceptable degree of precision. Call these  $\hat{CTE}^*$ , our best estimates of the true  $CTE$  values for the control portfolio.
3. Suppose  $\alpha_1$  is the lowest “level” for which a  $CTE$  estimate is required (e.g.,  $\alpha_1 = 0.70$ ) and  $M$  is the maximum number of scenarios over which the company is prepared to value the actual inforce portfolio (e.g.,  $M = 1000$ ).
4. Select  $M$  scenarios from the  $N \times (1 - \alpha_1)$  scenarios that comprise  $\hat{CTE}$ . Call these the “subset” scenarios. Clearly,  $N$  must be large enough so that  $N \times (1 - \alpha_1) \geq M$ . The subset could be randomly selected or the company could choose every  $w^{\text{th}}$  scenario. In any case, the average of the subset should closely approximate  $\hat{CTE}^*(\alpha_1)$  calculated from the full set.
5. For each  $\alpha_k$ , define an “adjustment factor”

$$\psi_k = \frac{\hat{CTE}^*(\alpha_k)}{\hat{CTE}'(\alpha_k)},$$

where  $\hat{CTE}'(\alpha_k)$  is estimated (for the control portfolio) from the subset scenarios. Note that  $\hat{CTE}'(\alpha_k)$  is calculated by averaging the “worst” (e.g., highest liabilities)

$$100 \times \left( \frac{1 - \alpha_k}{1 - \alpha_1} \right) \%$$

control results<sup>31</sup> from the  $M$  scenarios. By definition,  $\alpha_k \geq \alpha_1$  for all  $k$ .

6. Run the full (i.e., actual inforce) portfolio over the  $M$  subset scenarios. As stated previously, calculate  $\hat{CTE}''(\alpha_k)$  by averaging the “worst”

$$100 \times \left( \frac{1 - \alpha_k}{1 - \alpha_1} \right) \%$$

results from the  $M$  scenarios. The final  $CTE$  estimate for each  $\alpha_k$  is then obtained by multiplying by the corresponding adjustment factor. That is,

$$\hat{CTE}(\alpha_k) = \psi_k \times \hat{CTE}''(\alpha_k).$$

### REFERENCES

AAA LIFE CAPITAL ADEQUACY SUBCOMMITTEE. 2002. *Recommended Approach for Setting Regulatory Risk-Based Capital Requirements for Variable Products with Guarantees (Excluding Index Guarantees)*. American Academy of Actuaries. Online at [www.actuary.org/pdf/life/rbc\\_16dec02.pdf](http://www.actuary.org/pdf/life/rbc_16dec02.pdf).

ARTZNER, PHILIPPE. 1999. Application of Coherent Risk Measures to Capital Requirements In Insurance. *North American Actuarial Journal* 3(2): 11–25.

CIA COMMITTEE ON LIFE INSURANCE FINANCIAL REPORTING. 2001. *Guidance for the 2001 Valuation of Policy Liabilities of Life Insurers*. Canadian Institute of Actuaries. Online at [www.actuaries.ca/publications/2001/20180e.pdf](http://www.actuaries.ca/publications/2001/20180e.pdf).

<sup>29</sup> As a “control variate,” our experience suggests that it is often possible to capture the broad characteristics of large portfolios with as few as 50–100 representative “policies.”

<sup>30</sup> Typically,  $N \geq 10,000$ . For certain loss distributions, some applications might require  $N$  to be much higher.

<sup>31</sup> For example, if  $\alpha_1 = 0.7$ ,  $\alpha_k = 0.95$  and  $M = 1000$ , then we obtain  $\hat{CTE}'(95\%)$  by averaging the results from the  $M \times (1 - 0.95)/(1 - 0.70) = 167$  scenarios with the highest liabilities.

- CIA TASK FORCE ON SEGREGATED FUND INVESTMENT GUARANTEES. 2002. *Report of the CIA Task Force on Segregated Fund Investment Guarantees*. Canadian Institute of Actuaries. Online at [www.actuaries.ca/publications/2002/202012e.pdf](http://www.actuaries.ca/publications/2002/202012e.pdf).
- EMBRECHTS, PAUL, CLAUDIA KLÜPPELBERG, AND THOMAS MIKOSCH. 1997. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.
- HANCOCK, GEOFFREY H. 2003. PD-10: Variance of the CTE Estimator and its Application in CGAAP Valuation. *Canadian Institute of Actuaries Seminar for the Appointed Actuary*. Montréal.
- HARDY, MARY R. 2001. A Regime-Switching Model of Long-Term Stock Returns. *North American Actuarial Journal* 5(2): 41–53.
- . 2003. *Investment Guarantees—Modeling and Risk Management for Equity-Linked Life Insurance*. New York: John Wiley & Sons, Inc.
- JONES, BRUCE L., AND RICÁRDAS ZITIKIS. 2003. Empirical Estimation of Risk Measures and Related Quantities. *North American Actuarial Journal* 7(4): 44–54.
- MANISTRE, B. JOHN, AND GEOFFREY H. HANCOCK. 2002. Variance of the CTE Estimator. *37th Actuarial Research Conference*. Ontario, Canada: University of Waterloo.
- PANJER, HARRY H., AND GORDON E. WILLMOT. 1992. *Insurance Risk Models*. Schaumburg, Ill.: Society of Actuaries.
- STAUDTE, ROBERT. G., AND SIMON. J. SHEATHER 1990. *Robust Estimation and Testing*. New York: John Wiley & Sons, Ltd.
- WANG, SHAUN. 2002. *A Risk Measure that Goes Beyond Coherence*. SCOR Reinsurance Co.

*Discussions on this paper can be submitted until October 1, 2005. The authors reserve the right to reply to any discussion. Please see the Submission Guidelines for Authors on the inside back cover for instructions on the submission of discussions.*