QUANTIFYING AND CORRECTING THE BIAS IN ESTIMATED RISK MEASURES

BY

JOSEPH HYUN TAE KIM* AND MARY R. HARDY§

ABSTRACT

In this paper we explore the bias in the estimation of the Value at Risk and Conditional Tail Expectation risk measures using Monte Carlo simulation. We assess the use of bootstrap techniques to correct the bias for a number of different examples. In the case of the Conditional Tail Expectation, we show that application of the exact bootstrap can improve estimates, and we develop a practical guideline for assessing when to use the exact bootstrap.

1. INTRODUCTION

The focus of this paper is the estimation by simulation of the Value at Risk (VaR) and Conditional Tail Expectation (CTE) risk measures. Both risk measures are in common use for quantifying financial risk in insurance applications.

The use of simulation has proved problematic. Because insurers tend to use hundreds of thousands of model points (for example with a seriatim approach) each scenario run may take several computer minutes, with the result that a sample of only 1,000 projections may take over 20 hours of computation. However, the simulated estimators of the VaR and CTE risk measures are biased in general. Running more simulations to reduce the bias may not be feasible in light of the heavy cost, in terms of time and money, of each scenario. The bootstrap algorithm offers a method for improving the inference from a Monte Carlo sample, possibly more cheaply than increasing the sample size. In this paper we consider the magnitude and direction (positive or negative) of the bias arising from simulating the risk measures, and assess the advantages and disadvantages of using bootstrap techniques to correct the estimates for bias.

The two risk measures we consider have achieved widespread application in banking and insurance practice. The quantile risk measure — or Value at Risk, or VaR, — is used in banking for short term risks. The conditional tail

* Joseph Kim acknowledges the support in part of the Ph.D. Grant of the Society of Actuaries and the PGS-D2 grant of the Natural Sciences and Engineering Research Council of Canada.

§ Mary Hardy acknowledges the support of the Natural Sciences and Engineering Research Council of Canada.

Astin Bulletin 37(2), 365-386. doi: 10.2143/AST.37.2.2024072

expectation (CTE), also known as the tail conditional expectation, expected shortfall, or tailVaR, has been widely accepted in the insurance field, having been prescribed both by the Canadian Institute of Actuaries for Segregated Fund Contracts (CIA Segregated Funds Task Force (2002)) and by the American Academy of Actuaries in its 'C3 Phase 2' report (AAA Life Capital Adequacy Subcommittee (2005)).

Throughout the paper we assume F, the cumulative distribution function (c.d.f.) of random loss X, to be continuous. The quantile risk measure at the confidence level α, $0 < \alpha < 1$, is then simply

$$Q_\alpha(X) = F^{-1}(\alpha). \tag{1}$$

The CTE for a nonnegative loss random variable, at the confidence level α, $0 < \alpha < 1$, is defined as

$$CTE_\alpha(X) = \int_\alpha^1 F^{-1}(q)\,dq. \tag{2}$$

That is, the CTE is the expected value of the loss given that the loss falls in the upper $1-\alpha$ part of the distribution. If X is continuous at Q_α, the CTE is simply

$$E[X \mid X > Q_\alpha(X)] \tag{3}$$

In the sequel we often drop X from the notation and use CTE_α and Q_α for notational convenience. For the quantile risk measure we sometimes use $Q(\alpha)$ instead of Q_α when the confidence level is emphasized. The advantages of the CTE over the VaR risk measure are widely known. In particular, the CTE is coherent in the terms of Artzner *et al.* (1999)[1].

2. Estimating VaR and CTE using Monte Carlo

We assume that the Monte Carlo simulation generates an i.i.d. random sample $\mathbf{X} = (X_1, \ldots, X_n)$. The ordered sample is $(X_{(1)}, \ldots, X_{(n)})$.

For the sample quantile, $\widehat{Q}_\alpha$, there are several suggestions for estimators. The simplest candidate is $X_{(r)}$ where $(r-1)/n < \alpha \le r/n$. Slightly more sophisticated is an interpolation between $X_{(r-1)}$ and $X_{(r)}$ if $(r-1)/n < \alpha < r/n$. All of these estimators are biased, for finite samples, and are asymptotically unbiased.

Another adjustment is the use of $X_{(\alpha(n+1))}$, using some form of interpolation if $\alpha(n+1)$ is not an integer. For more discussion of these estimators, see for example, Hyndman and Fan (1996) and Klugman *et al.* (1998).

[1] The CTE has been claimed not to be coherent. The only problem arises if we try to use equation (3) with a random loss that is not continuous at Q_α. The original definition of the CTE in Wirch and Hardy (1999) specifically addresses this issue.

Given the same random sample $\mathbf{X} = (X_1, \ldots, X_n)$ the sample CTE estimate for confidence level α is generally taken as:

$$\widehat{CTE}_\alpha = \frac{1}{n(1-\alpha)} \sum_{i=[n\alpha]+1}^{n} X_{(i)}, \tag{4}$$

where the $X_{(i)}$ is the i-th ordered value of $\mathbf{X}$, and [ ] is the floor function.

These estimators for the CTE and VaR all take the form of a linear combination of order statistics which is commonly called the L-estimator (assuming any interpolation required is also linear).

Whenever the quantile estimator is biased, so is the CTE estimator, because

$$E[\widehat{CTE}_\alpha] = E_{\widehat{Q}_\alpha}[E_X(X|X > \widehat{Q}_\alpha)] \neq E_X[X|X > Q_\alpha] = CTE_\alpha. \tag{5}$$

The bias will tend to zero as the sample size gets larger since most suggested $\widehat{Q}_\alpha$s are consistent estimators of the true quantile, but will materially affect the accuracy of the CTE estimate for small sample sizes.

We illustrate the general problem with a simple analytic example.

Example 2.1. *Consider a uniform random variable* $X \sim U(0,1)$. *Suppose that we are interested in* α *such that* $n\alpha$ *is an integer. The true VaR here is* α. *The two simple estimators are* $X_{(n\alpha)}$ *and* $X_{(n\alpha+1)}$.

Now, for a sample size of n the expected value of r-th ordered value is given by $r/(n+1)$, $1 \le r \le n$.

If we set $\widehat{Q}_\alpha = X_{(n\alpha)}$, *the bias is* $-\alpha/(n+1)$. *On the other hand if* $\widehat{Q}_\alpha = X_{(n\alpha+1)}$ *the bias is* $(1-\alpha)/(n+1)$. *In this example the latter choice yields much smaller bias in absolute terms in the right tail region, where* α *is close to* 1.

This example shows that the direction and the magnitude of VaR bias depends on the choice of estimating function, the location of the quantile (for example, how close to the tail is it? and, which tail?), as well as the underlying distribution shape and the sample size. For a non-negative loss random variable, a negatively biased estimate of VaR could result in inadequacy of reserve or capital, and a positively biased estimate could cause inefficient use of capital.

Example 2.2. *For the same uniform distribution* $X \sim U(0,1)$ *the true CTE is* $(1+\alpha)/2$. *Assuming again that* $n\alpha$ *is an integer the expected value of the sample CTE estimate is*

$$\begin{aligned} E\left[\frac{1}{n(1-\alpha)} \sum_{i=n\alpha+1}^{n} X_{(i)}\right] &= \frac{1}{n(1-\alpha)} \sum_{i=n\alpha+1}^{n} E\left[X_{(i)}\right] \\ &= \frac{1}{n(1-\alpha)} \sum_{i=n\alpha+1}^{n} \frac{i}{n+1} \\ &= \frac{n(1+\alpha)+1}{2(n+1)}. \end{aligned}$$

So, the bias is $-\frac{\alpha}{2(n+1)}$, *meaning that the sample CTE tends to underestimate the true CTE in this example.*

We will see in the following section that the negative bias of the simulated CTE is a general observation.

3. Bias of sample estimates of VaR and CTE

We first focus on estimates using a single sample value $X_{(r)}$. We define the following two estimates and name them respectively the lower side estimate and the upper side estimate:

$$\widehat{Q}_L(\alpha) = X_{(r)}, \quad \text{if} \quad (r-1)/n < \alpha \le r/n \tag{6}$$

$$\widehat{Q}_U(\alpha) = X_{(r)}, \quad \text{if} \quad (r-1)/n \le \alpha < r/n \tag{7}$$

These are identical except when $n\alpha$ is integer. For example if $n = 100$ and $\alpha = 0.95$, $\widehat{Q}_L(0.95) = X_{(95)}$ whereas $\widehat{Q}_U(0.95) = X_{(96)}$.

As we saw in the Uniform example, a better estimator may lie between the low side and high side. There are many versions of estimators based on both the low side and high side sample values, as discussed in Hyndman and Fan (1996). Here we choose the one recommended by them:

$$\widehat{Q}_{HF}(\alpha) = (1-\gamma)\, X_{(g)} + \gamma X_{(g+1)}, \tag{8}$$

where $g = [(n + 1/3)\alpha + 1/3]$ and $\gamma = (n + 1/3)\alpha + 1/3 - g$. This is actually derived from the approximation of the incomplete beta function ratio and is known to be median unbiased of order $o(n^{-1/2})$. A slightly modified version of this estimate is also found in Klugman *et al.* (1998), termed the smoothed quantile estimate.

There are estimators for the quantile that use more than two values of $X_{(j)}$. Harrell and Davis (1982) proposed:

$$\widehat{Q}_{HD}(\alpha) = \sum_{j=1}^{n} w_j X_{(j)}, \tag{9}$$

where

$$w_j = \frac{\int_{(j-1)/n}^{j/n} t^{(n+1)\alpha - 1}(1-t)^{(n+1)(1-\alpha)-1}\,dt}{B\left[(n+1)\alpha, (n+1)(1-\alpha)\right]}$$

with $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$. This estimate is actually the exact bootstrap estimate of $E(X_{((n+1)\alpha)})$, even for non-integer $(n+1)\alpha$, as noted by Hutson and

Ernst (2000). We discuss the exact bootstrap in more detail in the following section. Mausser (2001) showed that $\widehat{Q}_{HD}(\alpha)$ performs better than $\widehat{Q}_{U}(\alpha)$ for the marginal VaR determination of some financial asset portfolios.

It is interesting to examine the bias of a single order statistic (that is, $X_{(r)}$ for integer r) against the true quantile when $n\alpha$ is integer. For example, is $E[X_{(95)}]$ larger or smaller than the true quantile $Q(0.95)$? Sometimes this type of question can be tackled using the quantile bounds. There are different bounds available for the expected value of one order statistic expressed as a function of parent population's quantile; see Section 4.4 of David (1981) and references therein. Those bounds are nonparametric but additional information, such as convexity of the distribution function, often leads to better bounds. A few useful bounds for an i.i.d. sample can be derived by the c-ordering equivalence as follows, where $F(x) = \Pr[X \le x]$ is the c.d.f. of X and n is the sample size.

1. $F(E[X_{(r)}]) \le (\ge) \dfrac{r}{n+1}$ if F is convex (concave)
2. $F(E[X_{(r)}]) \le (\ge) \dfrac{r-1}{n}$ if $1/F$ is concave (convex)
3. $F(E[X_{(r)}]) \le (\ge) \dfrac{r}{n}$ if $1/(1-F)$ is convex (concave)

It is possible for a distribution to satisfy more than one of the criteria above. These bounds can serve as an informal guideline if applied to the empirical distribution. For example if $1/F$ is concave, $E[X_{(n\alpha+1)}] \le Q(\alpha)$ and $E[X_{(n\alpha)}] \le Q(\alpha - 1/n)$, indicating that $\widehat{Q}_L(\alpha) = X_{(n\alpha)}$ is a bad choice since it makes the existing bias worse. For many common distributions, including the Normal, the Exponential, the Pareto, and the Gamma, for certain parameters, the $(n\alpha)$th order statistic can be shown to be negatively biased, or $E[X_{(n\alpha)}] \le Q(\alpha)$, by the convexity of $1/(1-F)$. While this underestimation of VaR is observed in many fat-tailed financial data there are examples with positive bias. Consider a special case of the Inverse Weibull, also known as the Fréchet distribution, with c.d.f.

$$F(x) = \exp[-x^{-a}], \quad a > 0, \ x > 0. \tag{10}$$

It can be shown that $1/(1-F)$ is concave in the right tail region for any $0 < a < 1$, thus $E[X_{(n\alpha)}] \ge Q(\alpha)$ when α is close to 1. This implies that the VaR based on historical data may actually exceed the true VaR for some very fat-tailed distributions. We finally note that the recent result on the bias of the VaR estimator by Inui *et al.* (2005) is equivalent to the first criterion above. Defining the VaR as a left side tail risk measure, they proved $E[X_{((n+1)\alpha)}]$ converges to $Q(\alpha)$ from below, or $E[X_{((n+1)\alpha)}] \le Q(\alpha)$ in the left tail area where F is convex. For nonnegative random variables whose c.d.f.s are concave, such as the Exponential distribution, the result does not hold.

Turning to the CTE, we have a result that the sample CTE estimate is always negatively biased. The CTE is a form of trimmed mean — where we trim the lower $n\alpha$ values of the sample of n losses. Rychlik (1998) showed that for any identically (but not necessarily independent) distributed random sample $(X_1, \ldots, X_n)$ the trimmed mean has the following upper bound:

$$E\left[\frac{1}{k+1-j}\sum_{i=j}^{k} X_{(i)}\right] \leq \frac{n}{n+1-j}\int_{(j-1)/n}^{1} Q(\alpha)\,d\alpha,$$

Where $Q(\alpha)$ is the α-quantile of the distribution of X. Plugging in $k = n$ and $j = n\alpha + 1$, assuming $n\alpha$ is an integer, gives the upper bound for the sample CTE

$$E\left[\frac{1}{n(1-\alpha)}\sum_{i=n\alpha+1}^{n} X_{(i)}\right] \leq \frac{1}{1-\alpha}\int_{\alpha}^{1} Q(x)\,dx = E\left[X \middle| X > Q(\alpha)\right].$$

Interestingly, if we set $j = k = n\alpha + 1$ the upper bound for expected value of the sample VaR, $\widehat{Q}_U(\alpha)$, becomes CTE_α, for an identically distributed sample. There is also a lower bound available, which is not considered here; see the reference for details.

We have shown in this section that common estimators for quantile and CTE risk measures are biased in general. One method for estimating and correcting for bias is through the bootstrap technique.

4. The bootstrap

4.1. Overview

The bootstrap methodology is particularly useful for non-parametric statistical inference. It has been widely applied by financial practitioners and actuaries. For a comprehensive treatment, see standard textbooks such as Efron and Tibshirani (1993), Davison and Hinkley (1997), or Hall (1992).

The core idea of the bootstrap is to create pseudo-samples by resampling (with replacement) from the original sample. The relationship of the pseudo-samples to the original sample replicates many features of the relationship of the original sample to its underlying distribution.

The basic procedure of the bootstrap can be sketched as follows. Suppose we have an i.i.d random sample $\mathbf{X} = (X_1, \ldots, X_n)$ from an unknown distribution, with c.d.f. F, and we are interested in a parameter $t(F)$ such as a quantile or CTE. Pseudo-samples are generated by sampling, with replacement, generating a new sample of the same size n as the original, from the empirical distribution function (e.d.f.) $\hat{F}$. The generated sample, denoted by $\mathbf{X}^*$, is called a bootstrap sample; the capital letter states that this too is a random sample, but from the e.d.f., indicated by superscript *. The statistic of interest using this generated sample then is denoted by $T^* = T(\mathbf{X}^*)$. We repeat the exercise R times for R

different bootstrap samples $\mathbf{X}_1^*, \ldots, \mathbf{X}_R^*$, each of size n. From each sample we generate the statistic of interest, that is, T_k^* from the k-th bootstrap sample, giving $T_j^* = T(\mathbf{X}_j^*)$, $j = 1, \ldots, R$. Finally the bootstrap estimate of the statistic T is given by

$$E[T|\hat{F}] = E^*[T^*] \approx \frac{1}{R}\sum_{j=1}^{R} T_j^*. \tag{11}$$

It is sometimes possible to compute $E[T|\hat{F}]$ analytically without actually performing the simulation. Usually, however, the resampling simulation, referred to the ordinary bootstrap (OB), is inevitable. In these cases the estimate is subject to sampling error. The difference between the true bootstrap estimate and the estimated one, called the resampling (simulation) error, decreases as the resampling size R gets larger. We denote $R^{-1}\sum T_j^*$ by $\bar{T}^*$ and call this the standard bootstrap estimator.

The unknown bias $E[T|F] - t(F)$ is approximated by its bootstrap estimate $B = E[T|\hat{F}] - t(\hat{F})$, thus the bootstrap bias estimate under R resamplings is

$$B_R = \bar{T}^* - t(\hat{F}) = \bar{T}^* - T. \tag{12}$$

Note that the bootstrap bias B_R converges to the true bias B as $R \to \infty$. If this is achieved the remaining uncertainty is only attributed to the original statistical error, i.e., to the fact that the empirical $\hat{F}$ does not perfectly represent the true F. The statistical error can be reduced when the sample size n gets larger or one has more information on F such as its parametric properties, neither of which might be feasible in practice.

We know that the estimator (e.g. for VaR or CTE) T is, in general, biased for finite sample sizes, so we could use a new bias-corrected estimator

$$t(\hat{F}) - B_R = 2T - \bar{T}^* \tag{13}$$

In practice, there is a trade-off between the improvement in the bias through using bias correction and the variability of the estimate in bootstrap. The problem is that the bias correction has its own variability. The efficiency of the estimator is often measured by the mean square error, or MSE, which is the sum of the squared bias and the variance of the estimator. The bias correction may reduce the contribution of the first term, but increase the second. Whether the resulting MSE will be smaller after bias correction depends on the underlying distribution as well as on the estimator itself; see for example Jeske and Sampath (2003). We will show some examples in the next section.

4.2. Bootstrapping L-estimators

For the quantile and CTE risk measures we can utilize results available on bootstrap estimation of L-estimators. Throughout this subsection we assume the statistic of interest, T, is an L-estimator, that is a linear combination of order statistics:

$$T = \sum_{i=1}^{n} c_i X_{(i)}. \tag{14}$$

It is evident that the sample VaR and the CTE are special cases of this form with appropriate selection of the coefficients c_i, $1 \le i \le n$. Hutson and Ernst (2000) derived a formula for the exact bootstrap (EB) mean and variance of any L-estimator. The bootstrap is exact in the sense that the resampling error is completely eliminated in the procedure; this is equivalent to the OB at $R = \infty$. We show here how to apply the exact bootstrap to the quantile measure.

Theorem 4.1. (Hutson and Ernst (2000)) *The exact bootstrap (EB) of the estimate of* $E(X_{(r)}|F)$, $1 \le r \le n$ *is*

$$E(X_{(r)}|\hat{F}) = \sum_{j=1}^{n} w_{j(r)} X_{(j)},$$

where

$$w_{j(r)} = r\binom{n}{r}\left[B\left(\frac{j}{n}; r, n-r+1\right) - B\left(\frac{j-1}{n}; r, n-r+1\right)\right],$$

and

$$B(x; a,b) = \int_0^x t^{a-1}(1-x)^{b-1} dt.$$

Some $w_{j(r)}$ values are presented in Figure 1.

The weights are spread around the sample estimate, with heavier weights around $X_{(r)}$ and gradually smaller weights for distant observations. The Harrell-Davis quantile estimator, introduced in Section 3, is equivalent to the EB of $E[X_{((n+1)\alpha)}|\hat{F}]$. Its weights can be compared to those of the EB at each order statistic. Figure 2 compares the 95% and 99% quantile weights. We can see from the figure that $\widehat{Q}_L^{EB} < \widehat{Q}_{HD} < \widehat{Q}_U^{EB}$ with $\widehat{Q}_{HD}$ close to $\widehat{Q}_U^{EB}$.

There are several advantages of the EB over the OB. The EB formula takes a simple analytic form, thus no simulations are involved. The simple form is easy to implement and significantly reduces the computing time compared to the OB, especially when sample sizes are big. Since no resampling error is involved in the EB, we expect the EB to be better than its OB counterpart in terms of variance on average. Finally the EB weights can be used for any samples with the same size because they are independent of the data. In the next section the EB is compared with the OB for both the VaR and the CTE using some parametric models.

Using Theorem 4.1 we now present several bootstrap-related quantities of the L-estimator in matrix form which is convenient in notation and useful for

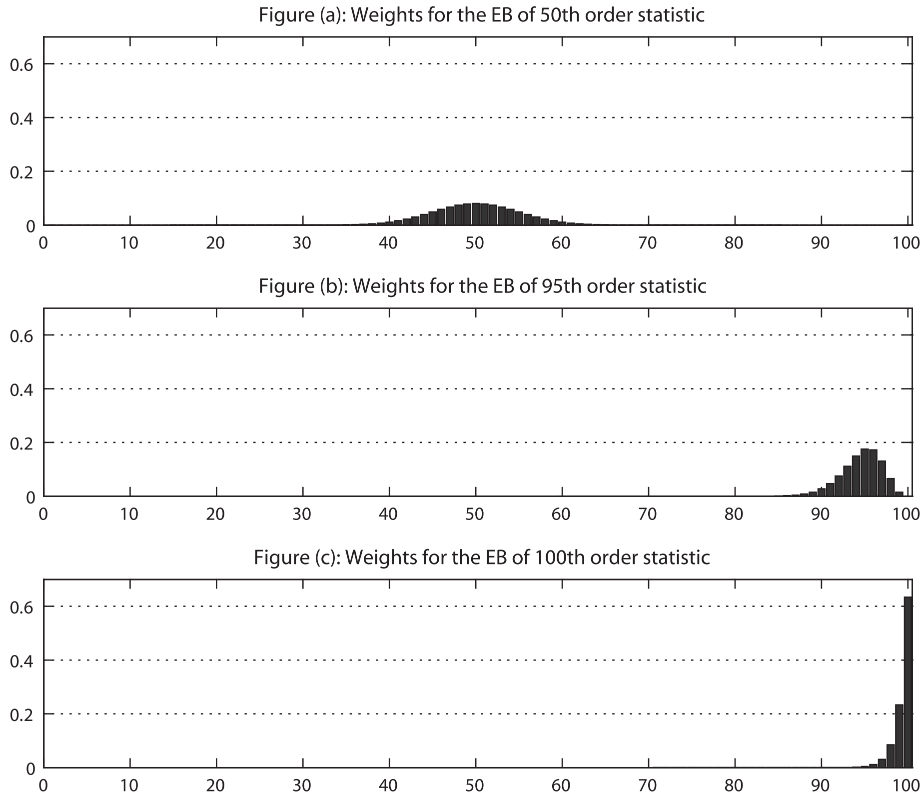

FIGURE 1: The weights of the EB for several order statistics when $n = 100$.

programming in matrix-based software such as Matlab. For a sample with size n, define $\mathbf{X}_{:n} = (X_{(1)}, \ldots, X_{(n)})'$ and $\mathbf{c}$ to be a column vector of size n. Then any L-estimator can be expressed as $T = \mathbf{c}'\mathbf{X}_{:n}$. For the CTE at confidence level α, for instance, we take

$$\mathbf{c} = (n(1-\alpha))^{-1}(0, \ldots, 0, 1, \ldots, 1)'$$

with zeros for the first $n\alpha$ elements, to get the sample CTE estimate $\mathbf{c}'\mathbf{X}_{:n}$. Therefore the EB of T, an L-estimator, is

$$E[T|\hat{F}] = E[\mathbf{c}'\mathbf{X}_{:n}|\hat{F}] = \mathbf{c}'E[\mathbf{X}_{:n}|\hat{F}] = \mathbf{c}'\mathbf{w}'\mathbf{X}_{:n}, \tag{15}$$

where the matrix $\mathbf{w} = \{w_{i(j)}\}_{i,j=1}^{n}$ comes from the EB weights for each element of $\mathbf{X}_{:\mathbf{n}}$. Now the EB bias estimate is expressed by

$$B = E[T|\hat{F}] - t(\hat{F}) = \mathbf{c}'\mathbf{w}'\mathbf{X}_{:n} - \mathbf{c}'\mathbf{X}_{:n} \tag{16}$$

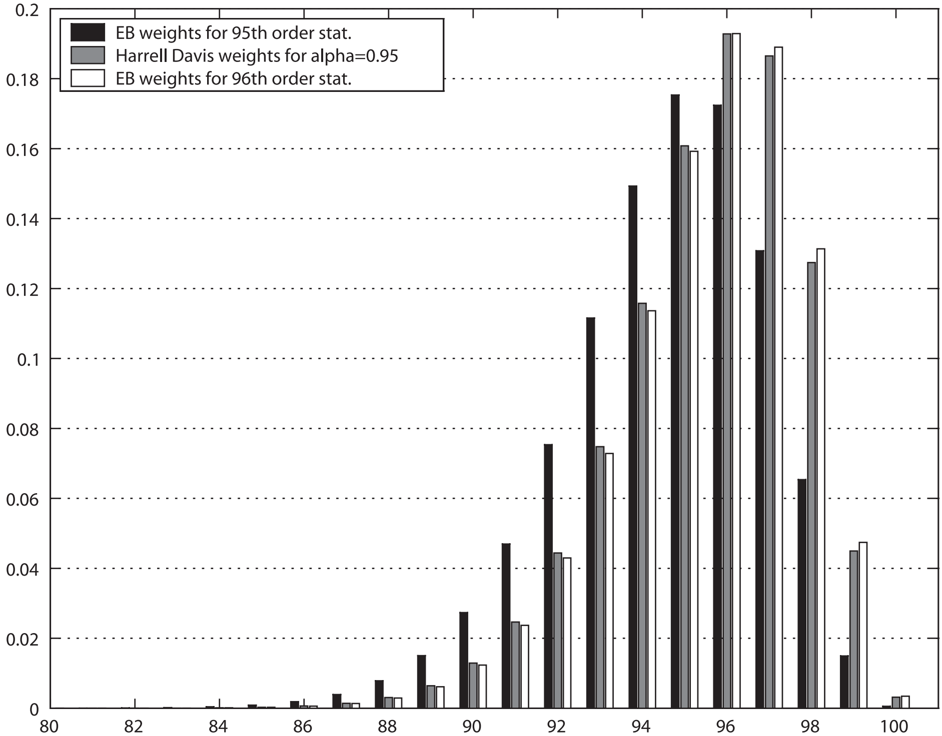

FIGURE 2: Weight comparison: Harrell-Davis (HD) vs. EB estimators when $n = 100$.

Thus the bias-corrected estimator defined in (13) then is

$$t(\hat{F}) - B = \mathbf{c}'\mathbf{X}_{:n} - (\mathbf{c}'\mathbf{w}'\mathbf{X}_{:n} - \mathbf{c}'\mathbf{X}_{:n}) = \mathbf{c}'(2\mathbf{I} - \mathbf{w}')\mathbf{X}_{:n}. \tag{17}$$

Even though we have an analytic expression for the EB bias it still is to be answered whether bias correction using the EB actually corrects the bias for the tail risk measures. The answer to this question is positive for the CTE as shown in the following theorem. See the Appendix for the proof.

Theorem 4.2. *For any given sample of size n, the empirical CTE estimator defined in (4) is always bigger than the EB of the CTE estimator, at any level α such that $n\alpha$ is an integer. That is, mathematically,*

$$\mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}} < \mathbf{c}'\mathbf{X}_{:\mathbf{n}},$$

for any given sample $\mathbf{X}$*, where* $\mathbf{c} = (n(1-\alpha))^{-1}(0, \ldots, 0, 1, \ldots, 1)'$ *with zeros for the first $n\alpha$ elements.*

Note that this result also holds for the ordinary bootstrap with sufficiently large resampling size R because the EB is the limit value of the ordinary bootstrap.

Coupled with the result of Rychlik (1998) the above theorem gives

$$E[\mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}}|F] < E[\mathbf{c}'\mathbf{X}_{:\mathbf{n}}|F] < CTE_\alpha(X),$$

for $\mathbf{c} = (n(1-\alpha))^{-1}(0, \ldots, 0, 1, \ldots, 1)'$ with zeros for the first $n\alpha$ elements. This implies that the bootstrap bias correction for CTE works in the right direction because the unknown bias $E[\mathbf{c}'\mathbf{X}_{:\mathbf{n}}|F] - CTE_\alpha(X)$ is estimated by $\mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}} - \mathbf{c}'\mathbf{X}_{:\mathbf{n}}$. Unfortunately there is no similar result for the VaR case.

5. Simulations

5.1. Examples

Three different examples are used to compare the performance of various estimators for the VaR and the CTE using Monte Carlo simulations; the empirical, the OB with resampling, the EB. For each bootstrap method the bias corrected estimator has also been computed. To assess the performance of different estimators we repeat these computations with different generated samples, leading to the MSE comparison.

The first example is a 10-year European put option with the price return based on the Lognormal (LN) distribution. The initial price of the asset is set at \$100, the strike price is \$180, and the risk free rate is assumed to be 0.5% per month effective. The LN parameters of the P-measure are $\mu = 0.00947$ and $\sigma = 0.04167$ which are derived from the monthly S&P 500 data during 1956-2001, as shown in Chapter 3 of Hardy (2003). Put options are often discussed in the cost of embedded investment guarantees such as in segregated funds and variable annuities, where the strike price represents the guaranteed payment at maturity to customers. The put option here can be said to be at the money from insurer's perspective because the expected level of the fund in 10 years under the risk neutral measure is around \$182. We focus on the VaR and the CTE of the put option liability at different confidence levels. We assume no hedging here, so these are risk measures for the naked liability.

In the second example we consider the identical put option except that the underlying asset follows the Regime Switching Log-Normal distribution with two regimes (RSLN2). See Hardy (2003) for details. The parameters are derived from the same S&P data: $\mu_1 = 0.0127$, $\mu_2 = -0.0162$, $\sigma_1 = 0.0351$, $\sigma_2 = 0.0691$, $p_{12} = 0.0468$, and $p_{21} = 0.3232$. Since the left tail of the RSLN2 is fatter than the LN the risk measure associated with the guarantee cost is known to be significantly greater under the former model than those under the latter, and this is exactly what we observe in our simulations.

The final model is a fat-tailed Pareto distribution which has been popular in connection with extreme value theory in financial risk management, and is also used in property and casualty applications. The parameters are $\beta = 10$ and $\xi = 0.2$ where the c.d.f. of Pareto is

$$F(x) = 1 - \left(\frac{\beta}{\beta + \xi x}\right)^{1/\xi}, \quad x > 0, \tag{18}$$

following the notation of Manistre and Hancock (2005). The mean and the variance of this distribution are respectively 12.5 and 260.42 under our parameter choice. Note that the Pareto is considered to be fatter tailed than the other two models.

For each model, we estimate the VaR using each of $\widehat{Q}_L$, $\widehat{Q}_U$, $\widehat{Q}_{HF}$, the bootstrap versions of the estimators using the ordinary bootstrap (OB) with 100 resamplings as in (11), the exact bootstrap (EB) as in (15), and the bias corrected estimators given in (13) and (17) respectively. This results in fifteen different estimators. Also we include $\widehat{Q}_{HD}$, for which there is no bootstrapping. For the CTE estimation we compare the empirical estimate, the OB and the EB without bias correction, and the OB and EB with bias correction giving a total of five estimators.

5.2. Estimating the 99% Quantile risk measure

We assume the actuary is using 200 or 1000 simulations to estimate the risk measure. To estimate the distribution of possible outcomes, we have repeated the 200 or 1000 simulations for a total of 20,000 different samples, for each liability model.

For each sample, we have calculated the lower quantile estimate, $\widehat{Q}_L$ from equation (6), the upper estimate $\widehat{Q}_U$ from equation (7), the HF estimate, $\widehat{Q}_{HF}$ from equation (8) and the HD estimate $\widehat{Q}_{HD}$ from equation (9). In addition, for the first three measures, for each sample, we have calculated revised estimates using the ordinary bootstrap and the exact bootstrap, before and after bias correction.

The resulting values give the bias and the root mean square error, rMSE, associated with each measure, averaged over the 20,000 samples. Estimated standard errors are also shown in percentage of the true value.

From these tables we note

1. The tail measure from the sample is quite inaccurate even for sample size 1000, which would be on the high side in many actuarial applications.
2. The overall accuracy, as measured by the rMSE is improved by using the exact bootstrap, without bias correction.
3. Bias correction may increase the bias and substantially increase the rMSE — and in fact does so for the lognormal and RSLN models, for the $\widehat{Q}_U$ and $\widehat{Q}_{HF}$. The usefulness of the bootstrap bias correction is limited because of the non-smoothness of the estimator, $\widehat{Q}_U = X_{(0.99n)}$, and further because

TABLE 1

99% QUANTILE ESTIMATORS FOR THE LOGNORMAL EXAMPLE

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	StD (s.e.)	rMSE	Bias (s.e.)	StD (s.e.)	rMSE
$\hat{Q}_L$	−7.59%(0.12%)	16.70%(1.11%)	18.34%	−1.58%(0.06%)	7.91%(0.25%)	8.07%
$\hat{Q}_L^{OB}$	−9.34%(0.10%)	14.66%(0.85%)	17.38%	−1.94%(0.05%)	7.37%(0.22%)	7.62%
$\hat{Q}_L^{OB.bc}$	−5.84%(0.14%)	20.44%(1.66%)	21.26%	−1.21%(0.06%)	9.01%(0.32%)	9.09%
$\hat{Q}_L^{EB}$	−9.33%(0.10%)	14.55%(0.84%)	17.28%	−1.94%(0.05%)	7.32%(0.21%)	7.57%
$\hat{Q}_L^{EB.bc}$	−5.85%(0.14%)	20.36%(1.65%)	21.18%	−1.22%(0.06%)	8.96%(0.32%)	9.04%
$\hat{Q}_U$	4.69%(0.13%)	18.25%(1.32%)	18.84%	0.98%(0.06%)	8.05%(0.26%)	8.11%
$\hat{Q}_U^{OB}$	1.84%(0.11%)	15.75%(0.99%)	15.86%	0.58%(0.05%)	7.50%(0.22%)	7.52%
$\hat{Q}_U^{OB.bc}$	7.54%(0.16%)	22.81%(2.07%)	24.02%	1.38%(0.06%)	9.19%(0.34%)	9.29%
$\hat{Q}_U^{EB}$	1.84%(0.11%)	15.63%(0.97%)	15.74%	0.59%(0.05%)	7.45%(0.22%)	7.47%
$\hat{Q}_U^{EB.bc}$	7.55%(0.16%)	22.71%(2.05%)	23.93%	1.37%(0.06%)	9.14%(0.33%)	9.24%
$\hat{Q}_{HF}$	0.56%(0.12%)	16.93%(1.14%)	16.94%	0.12%(0.06%)	7.92%(0.25%)	7.92%
$\hat{Q}_{HF}^{OB}$	−1.92%(0.11%)	15.23%(0.92%)	15.35%	−0.27%(0.05%)	7.45%(0.22%)	7.45%
$\hat{Q}_{HF}^{OB.bc}$	3.04%(0.14%)	19.83%(1.56%)	20.06%	0.50%(0.06%)	8.84%(0.31%)	8.86%
$\hat{Q}_{HF}^{EB}$	−1.92%(0.11%)	15.13%(0.91%)	15.25%	−0.26%(0.05%)	7.40%(0.22%)	7.40%
$\hat{Q}_{HF}^{EB.bc}$	3.04%(0.14%)	19.74%(1.55%)	19.98%	0.50%(0.06%)	8.80%(0.31%)	8.81%
$\hat{Q}_{HD}$	1.72%(0.11%)	15.61%(0.97%)	15.71%	0.56%(0.05%)	7.45%(0.22%)	7.47%

True Value **39.7202**

we are near the bounds of the e.d.f. sample space. To illustrate more clearly, for the 99.5% quantile, $\widehat{Q}_U$ would be the maximum value from the sample, which we expect to be positively biased for the models under consideration. But a bootstrap estimate of a sample maximum could never be positively biased.

4. Even in cases where the bias correction reduces the average bias, the resulting increase in the standard error of the estimator for these examples leads to a bigger rMSE.
5. The EB is always more efficient than the OB as we would expect, even though the improvement is marginal. The improvement is not statistically significant, but the systematic reduction in error is due to the elimination of bootstrap sampling volatility, thus it is strongly recommended to use the EB instead of the OB whenever possible.
6. The $\widehat{Q}_{HF}$ and $\widehat{Q}_{HF}^{EB}$ estimators perform well in all cases; if the estimator efficiency is important, this may be a good default selection.

TABLE 2

99% QUANTILE ESTIMATORS FOR THE RSLN2 EXAMPLE

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	StD (s.e.)	rMSE	Bias (s.e.)	StD (s.e.)	rMSE
$\hat{Q}_L$	–6.06%(0.09%)	13.25%(0.91%)	14.57%	–1.27%(0.04%)	6.29%(0.21%)	6.42%
$\hat{Q}_L^{OB}$	–7.58%(0.08%)	11.61%(0.70%)	13.87%	–1.57%(0.04%)	5.84%(0.18%)	6.05%
$\hat{Q}_L^{OB.bc}$	–4.55%(0.11%)	16.21%(1.36%)	16.83%	–0.97%(0.05%)	7.18%(0.27%)	7.24%
$\hat{Q}_L^{EB}$	–7.58%(0.08%)	11.53%(0.69%)	13.80%	–1.57%(0.04%)	5.81%(0.17%)	6.02%
$\hat{Q}_L^{EB.bc}$	–4.54%(0.11%)	16.14%(1.35%)	16.77%	–0.96%(0.05%)	7.15%(0.26%)	7.21%
$\hat{Q}_U$	3.50% (0.1%)	14.27%(1.06%)	14.69%	0.74%(0.05%)	6.41%(0.21%)	6.45%
$\hat{Q}_U^{OB}$	1.21%(0.09%)	12.33%(0.79%)	12.39%	0.42%(0.04%)	5.92%(0.18%)	5.94%
$\hat{Q}_U^{OB.bc}$	5.79%(0.13%)	17.76%(1.64%)	18.68%	1.05%(0.05%)	7.35%(0.28%)	7.42%
$\hat{Q}_U^{EB}$	1.22%(0.09%)	12.25%(0.78%)	12.31%	0.42%(0.04%)	5.89%(0.18%)	5.91%
$\hat{Q}_U^{EB.bc}$	5.78%(0.12%)	17.68%(1.62%)	18.6%	1.05%(0.05%)	7.31%(0.28%)	7.39%
$\hat{Q}_{HF}$	0.28%(0.09%)	13.31%(0.92%)	13.31%	0.06%(0.04%)	6.3%(0.21%)	6.30%
$\hat{Q}_{HF}^{OB}$	–1.75%(0.08%)	11.98%(0.74%)	12.1%	–0.25%(0.04%)	5.89%(0.18%)	5.89%
$\hat{Q}_{HF}^{OB.bc}$	2.31%(0.11%)	15.57%(1.26%)	15.74%	0.37%(0.05%)	7.07%(0.26%)	7.08%
$\hat{Q}_{HF}^{EB}$	–1.75%(0.08%)	11.91%(0.74%)	12.04%	–0.25%(0.04%)	5.86%(0.18%)	5.86%
$\hat{Q}_{HF}^{EB.bc}$	2.31%(0.11%)	15.49%(1.25%)	15.66%	0.37%(0.05%)	7.04%(0.26%)	7.05%
$\hat{Q}_{HD}$	1.12%(0.09%)	12.24%(0.78%)	12.29%	0.40%(0.04%)	5.89%(0.18%)	5.90%

True Value **51.8618**

7. Note that $\widehat{Q}_{HD}$ performs well and often ranks the second best followed by $\widehat{Q}_{HF}$, but its performance deteriorates when $\widehat{Q}_{HF}^{EB}$ fails to be the best. As we expected $\widehat{Q}_L^{EB} < \widehat{Q}_{HD} < \widehat{Q}_U^{EB}$, and $\widehat{Q}_{HD}$ is close to $\widehat{Q}_U^{EB}$ throughout the simulations.

5.3. Estimating the 95% CTE

Following a similar process to the quantile results above, we simulated 20,000 samples of 200 values, and 20,000 samples each with 1,000 values. For each sample, we estimated the 95% CTE directly from the sample, (as the mean of the largest 5% of simulated loss values) and then again using the ordinary bootstrap (with 100 bootstrap replications) and the exact bootstrap, without and with bias correction. The results are then averaged over the 10,000 simulations, and the final averages are shown in Tables 4, 5, and 6.

1. We notice that, as expected from Section 3 the sample estimates of the CTE are negatively biased.

TABLE 3

99% QUANTILE ESTIMATORS FOR THE PARETO EXAMPLE

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	StD (s.e.)	rMSE	Bias (s.e.)	StD (s.e.)	rMSE
$\hat{Q}_L$	–5.86%(0.15%)	21.01%(3.34%)	21.81%	–1.36%(0.07%)	10.24%(0.79%)	10.33%
$\hat{Q}_L^{OB}$	–4.01%(0.14%)	20.03%(3.03%)	20.43%	–1.02%(0.07%)	9.70%(0.71%)	9.76%
$\hat{Q}_L^{OB.bc}$	–7.71%(0.18%)	26.03%(5.12%)	27.15%	–1.71%(0.08%)	11.58%(1.01%)	11.71%
$\hat{Q}_L^{EB}$	–3.99%(0.14%)	19.84%(2.97%)	20.23%	–1.01%(0.07%)	9.64% (0.7%)	9.70%
$\hat{Q}_L^{EB.bc}$	–7.73%(0.18%)	25.82%(5.04%)	26.95%	–1.71%(0.08%)	11.51%(1%)	11.64%
$\hat{Q}_U$	11.90%(0.22%)	30.65%(7.10%)	32.88%	1.99%(0.08%)	11.03%(0.92%)	11.2%
$\hat{Q}_U^{OB}$	13.38%(0.21%)	29.85%(6.74%)	32.71%	2.40%(0.07%)	10.42%(0.82%)	10.69%
$\hat{Q}_U^{OB.bc}$	10.43%(0.28%)	39.61%(11.86%)	40.96%	1.59%(0.09%)	12.53%(1.19%)	12.63%
$\hat{Q}_U^{EB}$	13.40%(0.21%)	29.63%(6.64%)	32.52%	2.41%(0.07%)	10.35%(0.81%)	10.63%
$\hat{Q}_U^{EB.bc}$	10.40%(0.28%)	39.40%(11.74%)	40.75%	1.58%(0.09%)	12.46%(1.17%)	12.56%
$\hat{Q}_{HF}$	5.92%(0.18%)	26.12%(5.16%)	26.78%	0.86%(0.08%)	10.64%(0.86%)	10.68%
$\hat{Q}_{HF}^{OB}$	7.52%(0.19%)	26.20%(5.19%)	27.26%	1.25%(0.07%)	10.17%(0.78%)	10.24%
$\hat{Q}_{HF}^{OB.bc}$	4.32%(0.22%)	31.48%(7.49%)	31.77%	0.48%(0.08%)	11.83%(1.06%)	11.84%
$\hat{Q}_{HF}^{EB}$	7.55%(0.18%)	26.02%(5.12%)	27.09%	1.25%(0.07%)	10.10%(0.77%)	10.18%
$\hat{Q}_{HF}^{EB.bc}$	4.30%(0.22%)	31.29%(7.40%)	31.58%	0.47%(0.08%)	11.76%(1.05%)	11.77%
$\hat{Q}_{HD}$	13.19%(0.21%)	29.49%(6.57%)	32.30%	2.37%(0.07%)	10.35%(0.81%)	10.61%

True Value **75.594**

2. Unlike the Quantile examples, the bias correction for the CTE does reduce the bias in all cases, on average. However, in some cases the reduction in bias is outweighed by the resulting increase in variance, to give a slightly higher overall rMSE.
3. The case for using the bootstrap is not clear; the sample alone gives about the same accuracy as the bootstrapped variations, on average. In general, we find the EB with bias correction offers a very similar rMSE to the standard estimator, but with a smaller bias, which may be preferable.
4. Considering that in practice one has a single sample, the bootstrap might be the only sensible method to estimate the bias with no reference to the true distribution.

As a guideline as to whether to apply the bias correction, Efron and Tibshirani (1993) suggest that the ratio of the bootstrap bias estimate to the bootstrap standard error should be considered. If the ratio is bigger than 0.25 the bias correction is worth using.

TABLE 4

95% CTE ESTIMATORS FOR THE LOGNORMAL EXAMPLE

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	StD (s.e.)	rMSE	Bias (s.e.)	StD (s.e.)	rMSE
$\widehat{CTE}$	–2.68%(0.12%)	16.89%(0.89%)	17.10%	–0.52%(0.05%)	7.42%(0.17%)	7.44%
$\widehat{CTE}^{OB}$	–5.38%(0.12%)	16.62%(0.86%)	17.47%	–1.06%(0.05%)	7.43%(0.17%)	7.50%
$\widehat{CTE}^{OB.bc}$	0.01%(0.12%)	17.36%(0.94%)	17.36%	0.02%(0.05%)	7.50%(0.18%)	7.50%
$\widehat{CTE}^{EB}$	–5.37%(0.12%)	16.55%(0.86%)	17.4%	–1.06%(0.05%)	7.39%(0.17%)	7.47%
$\widehat{CTE}^{EB.bc}$	0%(0.12%)	17.27%(0.93%)	17.27%	0.02%(0.05%)	7.46%(0.17%)	7.46%

True Value **31.2552**

TABLE 5

95% CTE ESTIMATORS FOR THE RSLN2 EXAMPLE

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	StD (s.e.)	rMSE	Bias (s.e.)	StD (s.e.)	rMSE
$\widehat{CTE}$	–2.08%(0.09%)	12.68%(0.69%)	12.85%	–0.40%(0.04%)	5.62%(0.14%)	5.64%
$\widehat{CTE}^{OB}$	–4.17%(0.09%)	12.51%(0.67%)	13.18%	–0.83%(0.04%)	5.62%(0.14%)	5.68%
$\widehat{CTE}^{OB.bc}$	0.01%(0.09%)	13.01%(0.73%)	13.01%	0.02%(0.04%)	5.69%(0.14%)	5.69%
$\widehat{CTE}^{EB}$	–4.16%(0.09%)	12.44%(0.67%)	13.12%	–0.82%(0.04%)	5.60%(0.13%)	5.66%
$\widehat{CTE}^{EB.bc}$	0.01%(0.09%)	12.95%(0.72%)	12.95%	0.01%(0.04%)	5.65%(0.14%)	5.65%

True Value **42.9634**

TABLE 6

95% CTE ESTIMATORS FOR THE PARETO EXAMPLE

	Sample Size 200			Sample Size 1000		
	Bias (s.e.)	StD (s.e.)	rMSE	Bias (s.e.)	StD (s.e.)	rMSE
$\widehat{CTE}$	–1.32%(0.13%)	17.99%(2.06%)	18.03%	–0.33%(0.06%)	8.10%(0.42%)	8.11%
$\widehat{CTE}^{OB}$	–2.71%(0.13%)	17.75%(2.01%)	17.95%	–0.6%(0.06%)	8.11%(0.42%)	8.13%
$\widehat{CTE}^{OB.bc}$	0.08%(0.13%)	18.41%(2.16%)	18.41%	–0.06%(0.06%)	8.17%(0.43%)	8.17%
$\widehat{CTE}^{EB}$	–2.69%(0.12%)	17.67%(1.99%)	17.88%	–0.60%(0.06%)	8.07%(0.42%)	8.09%
$\widehat{CTE}^{EB.bc}$	0.06%(0.13%)	18.31%(2.14%)	18.31%	–0.06%(0.06%)	8.13%(0.42%)	8.13%

True Value **63.7853**

If the bias is not large relative to the standard error, it may be worth using the EB estimate even without bias correction rather than the empirical estimator. We see from the tables that in some cases the EB estimate is preferred, at other times the empirical estimator has smaller rMSE. An interesting question is how an actuary with a single sample from an unknown underlying distribution should decide whether applying the exact bootstrap will improve the estimator of the CTE or not. To help with this decision we have developed a guideline which is described in the following section.

6. EB or empirical CTE estimator? A practical test

In the previous section the true risk measure values were available in comparing different estimators' performances, but in practice one would have only a single sample with no information on the true value. We propose a practical guideline for the CTE estimation that can be used to select a better estimator in practice.

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two different estimators of an unknown parameter θ. Assume, say, $E[\hat{\theta}_1] < E[\hat{\theta}_2]$. For $\hat{\theta}_1$ to have smaller MSE than $\hat{\theta}_2$,

$$E[\theta - \hat{\theta}_1]^2 < E[\theta - \hat{\theta}_2]^2$$

which rearranges to:

$$\theta < \frac{1}{2}\left[\frac{\operatorname{Var}\left(\hat{\theta}_2\right) - \operatorname{Var}\left(\hat{\theta}_1\right)}{E\left[\hat{\theta}_2\right] - E\left[\hat{\theta}_1\right]} + E\left[\hat{\theta}_2\right] + E\left[\hat{\theta}_1\right]\right]. \tag{19}$$

Now, we will use this to compare the empirical CTE estimate, $\hat{\theta}_2 = \mathbf{c}'\mathbf{X}_{:\mathbf{n}}$ with the empirical EB estimator, $\hat{\theta}_1 = \mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}}$. From Theorem 4.2 we have $E[\hat{\theta}_2] - E[\hat{\theta}_1] > 0$.

The right hand side of inequality 19 then becomes:

$$\frac{1}{2}\left[\frac{\mathbf{c}'\boldsymbol{\Sigma}_{:\mathbf{n}}\mathbf{c} - \mathbf{c}'\mathbf{w}'\boldsymbol{\Sigma}_{:\mathbf{n}}\mathbf{w}\mathbf{c}}{\mathbf{c}'\mathbf{E}[\mathbf{X}_{:\mathbf{n}}] - \mathbf{c}'\mathbf{w}'\mathbf{E}[\mathbf{X}_{:\mathbf{n}}]} + \mathbf{c}'\mathbf{E}[\mathbf{X}_{:\mathbf{n}}] + \mathbf{c}'\mathbf{w}'\mathbf{E}[\mathbf{X}_{:\mathbf{n}}]\right]$$
$$= \frac{1}{2}\left[\frac{\mathbf{c}'\left(\boldsymbol{\Sigma}_{:\mathbf{n}} - \mathbf{w}'\boldsymbol{\Sigma}_{:\mathbf{n}}\mathbf{w}\right)\mathbf{c}}{\mathbf{c}'(\mathbf{I} - \mathbf{w}')\mathbf{E}[\mathbf{X}_{:\mathbf{n}}]} + \mathbf{c}'(\mathbf{I} + \mathbf{w}')\mathbf{E}[\mathbf{X}_{:\mathbf{n}}]\right],$$

where $\boldsymbol{\Sigma}_{:\mathbf{n}}$ is the covariance matrix with $Cov(X_{(i)}, X_{(j)})$ for each element. This formula again is approximated by plugging in the bootstrap estimator to give

$$\eta \equiv \frac{1}{2}\left|\frac{\mathbf{c}'\left(\hat{\Sigma}_{:\mathbf{n}} - \mathbf{w}'\hat{\Sigma}_{:\mathbf{n}}\mathbf{w}\right)\mathbf{c}}{\mathbf{c}'(\mathbf{I} - \mathbf{w}')\mathbf{w}'\mathbf{X}_{:\mathbf{n}}} + \mathbf{c}'(\mathbf{I} + \mathbf{w}')\mathbf{w}'\mathbf{X}_{:\mathbf{n}}\right|. \tag{20}$$

TABLE 7

APPLICATION OF MSE TEST TO PARETO AND RSLN EXAMPLE, 99% CTEs.

CTE 99% in the Lognormal model CTE = 47.7281; sample size n = 200				
Method	Estimate	Bias	Std	rMSE
Empirical	45.4203	– 2.3078	7.2700	7.6276
EB	42.9971	– 4.7310	6.6571	8.1669
Mixed	45.3762	– 2.3519	7.2796	7.6501
CTE 99% in the RSLN put model CTE = 59.9989; sample size n = 200				
Method	Estimate	Bias	Std	rMSE
Empirical	57.6421	– 2.3568	7.2434	7.6172
EB	55.1389	– 4.8600	6.6605	8.2451
Mixed	57.5984	– 2.4005	7.2513	7.6383
CTE 99% in the Pareto model CTE = 106.993; sample size n = 200				
Method	Estimate	Bias	Std	rMSE
Empirical	100.6815	– 6.3114	33.7357	34.3210
EB	93.9402	– 13.0527	29.2711	32.0495
Mixed	99.6372	– 7.3557	32.2089	33.0382

The bootstrap estimate of the covariance matrix[2] can be obtained through the OB, for each given sample, with each element $Cov(X_{(i)}, X_{(j)} | \hat{F})$. So, the EB estimator, θ_1, will be more efficient if the true (unknown) $\theta < \eta$, otherwise the empirical estimator should be used. We can substitute $\mathbf{c}'\mathbf{X}_{:\mathbf{n}} + B$ for θ to give an approximate rule of thumb, where B is the estimated bias,

$$B = \mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}} - \mathbf{c}'\mathbf{X}_{:\mathbf{n}}.$$

That is, use the EB estimate if

$$\mathbf{c}'\mathbf{X}_{:\mathbf{n}} + B < \eta. \tag{21}$$

Otherwise, use the empirical estimate without applying the exact bootstrap.

[2] Hutson and Ernst (2000) also provides the EB estimate of the covariance matrix analytically, but its computation increases exponentially as sample size gets larger. Even for n = 400, computing the EB variance requires prohibitive time and storage.

To illustrate this, we generate 10,000 samples, each with 200 values of the same three models to estimate 99% CTEs. For each sample, we apply the test in equation (21). Resampling size of 999 was used to estimate the covariance matrix for each sample. If the inequality is satisfied, we use the EB estimate. If it was not, we use the empirical estimate. The average outcome is labeled 'Mixed' in Table 7. In general, in practice, we assume the underlying model is unknown, and only one sample is available. The tables illustrate that using the test gives an average rMSE that is near the lower of the two, successfully identifying in the majority of cases whether the EB or empirical estimate is preferred. In the Pareto case, the exact bootstrap offers a lower MSE, and in the LN and the RSLN case the empirical estimates are better.

7. Conclusion

We investigated the bias of estimates of two risk measures, the quantile and the CTE, in finite samples. For the quantile, different estimators are compared with bootstrapped and bias corrected bootstrapped estimators. Simulations show that the exact bootstrap has definite advantages over the ordinary resampled bootstrap. The bootstrap bias correction however should not be applied to tail quantiles. The exact bootstrap offers a reasonably efficient estimator in many cases.

For the CTE we found the sample estimate is always negatively biased and the bias correction works reasonably well, though an increase in the variance may decrease the efficiency of the estimator with bias correction. Finally, we propose an algorithm to help determine whether the exact bootstrap estimator will be more efficient than the ordinary empirical estimate.

Before closing the paper, we note that the bias is not the only source of error, even though it has been focused on in this paper for small sample sizes. The variance of the estimated tail measures is relatively larger and decreases slower than the bias, as shown in the examples. This suggests that practitioners should also look at the magnitude of the estimated variance, e.g., using the bootstrap, for an accurate tail measure estimation. Since the bootstrap is based on the sample in hand, the sampling error of the original data cannot be cured through the bootstrap, especially when the quantity of interest lies in the tail region where events are rare.

Acknowledgements

The authors acknowledge very helpful comments from Jed Frees, Paul Marriott and two anonymous referees.

References

AAA Life Capital Adequacy Subcommittee (2005) Recommended approach for setting regulatory risk-based capital requirements for variable annuities and similar products. Report, American Academy of Actuaries, Boston, MA.

ARTZNER, P., DELBAEN, F., EBER, J.M. and HEATH, D. (1999) Coherent measure of risk. *Mathematical Finance*, 203-228.

CIA SEGREGATED FUNDS TASK FORCE (2002) Report of the CIA task force on segregated fund investment guarantees. Technical report, Canadian Institute of Actuaries, Ottawa, ON.

DAVID, H.A. (1981) *Order statistics*. John Wiley & Sons, New York, 2nd edition.

DAVISON, A.C. and HINKLEY, D.V. (1997) *Bootstrap methods and their application*. Cambridge University Press, New York.

EFRON, B. and TIBSHIRANI, R.J. (1993) *An introduction to the bootstrap*. Chapman & Hall, New York.

HALL, P. (1992) *The Bootstap and Edgeworth Expansion*. Springer Series in Statistics. Springer-Verlag, New York.

HARDY, M.R. (2003) *Investment Guarantees: Modeling and Risk Management for Equity-Linked Life Insurance*. John Wiley & Sons, New York.

HARRELL, F.E. and DAVIS, C.E. (1982) A new distribution-free quantile estimator. *Biometrika*, **69(3)**, 635-640.

HUTSON, A.D. and ERNST, M.D. (2000) The exact bootstrap mean and variance of an L-estimator. *Journal of Royal Statistical Society: Series B*, **62**, 89-94.

HYNDMAN, R.J. and FAN, Y. (1996) Sample quantiles in statistical packages. *The American Statistician*, **50(4)**, 361-365.

INUI, K., KIJIMA, M. and KITANO, A. (2005) VaR is subject to a significant positive bias. *Statistics & Probability Letters*, **72**, 299-311.

JESKE, D.R. and SAMPATH, A. (2003) A real example that illustrates interesting properties of boostrap bias correction. *The American Statistician*, **57(1)**, 62-65.

JOHNSON, N.L., KOTZ, S. and KEMP, A.W. (1992) *Univariate Discrete Distribution*. John Wiley & Sons, New York, 2nd edition.

KLUGMAN, S., PANJER, H. and WILLMOT, G. (1998) *Loss Models*. John Wiley, New York.

MANISTRE, B.J. and HANCOCK, G.H. (2005) Variance of the CTE estimator. *North American Actuarial Journal*, **9(2)**, 129-156.

MAUSSER, H. (2001) Calculating quantile-based risk analytics with L-estimators. *Algo Research Quarterly*, **4(4)**, 33-47.

RYCHLIK, T. (1998) Bounds for expectations of L-estimates. In N. Balakrishnan and C.R. Rao, editors, *Handbook of Statistics*, volume 16, Amsterdam. Elsevier.

WIRCH, J.L. and HARDY, M.R. (1999) A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, **25**, 337-347.

JOSEPH HYUN TAE KIM
Statistics and Actuarial Science
University of Waterloo
Waterloo, ON, Canada

Appendix: Proof of Theorem 4.2

The goal is to prove

$$\mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}} < \mathbf{c}'\mathbf{X}_{:\mathbf{n}},$$

for $\mathbf{c} = (n(1-\alpha))^{-1}(0, \ldots, 0, 1, \ldots, 1)'$ with zeros for the first $n\alpha$ elements. To start with, let us show that the weight matrix $\mathbf{w} = \{w_{i(j)}\}_{i,j=1}^{n}$ is doubly stochastic, meaning that the sum over each row and each column equals one. From Theorem 4.1

$$\begin{aligned}
w_{j(r)} &= r\binom{n}{r}\left[B\left(\frac{j}{n}; r, n-r+1\right) - B\left(\frac{j-1}{n}; r, n-r+1\right)\right] \\
&= \frac{1}{B(r, n-r+1)}\left[B\left(\frac{j}{n}; r, n-r+1\right) - B\left(\frac{j-1}{n}; r, n-r+1\right)\right] \\
&= \sum_{k=r}^{n}\binom{n}{k}\left(\frac{j}{n}\right)^{k}\left(1-\frac{j}{n}\right)^{n-k} - \sum_{k=r}^{n}\binom{n}{k}\left(\frac{j-1}{n}\right)^{k}\left(1-\frac{j-1}{n}\right)^{n-k}
\end{aligned}$$

The last line is a well known characteristic of the incomplete beta function ratio; see e.g., Johnson *et al.* (1992). If we define two binomial random variables $Y_1 \sim Bin(n, \frac{j}{n})$ and $Y_2 \sim Bin(n, \frac{j-1}{n})$, the last expression becomes $Pr(Y_1 \geq r) - Pr(Y_2 \geq r)$. Thus the sum of j-th row elements of $\mathbf{w}$ is

$$\begin{aligned}
\sum_{r=1}^{n} w_{j(r)} &= \sum_{r=1}^{n} Pr[Y_1 \geq r] - \sum_{r=1}^{n} Pr[Y_2 \geq r] \\
&= E[Y_1] - E[Y_2] = n\frac{j}{n} - n\frac{j-1}{n} = 1.
\end{aligned}$$

The sum over each column equals one by definition of the weights. Now put $\mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}} = \sum_{i=1}^{n} b_i X_{(i)}$ where $b_i = (n(1-\alpha))^{-1}\sum_{j=n\alpha+1}^{n} w_{i(j)}$. Note that for each i we have

$$0 < b_i < (n(1-\alpha))^{-1} \text{ with } \sum_{i=1}^{n} b_i = 1$$

because $\mathbf{w}$ is doubly stochastic. Thus

$$\begin{aligned}
\mathbf{c}'\mathbf{X}_{:\mathbf{n}} - \mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}} &= \sum_{i=n\alpha+1}^{n}\left(n(1-\alpha)\right)^{-1} X_{(i)} - \sum_{i=1}^{n} b_i X_{(i)} \\
&= \sum_{i=n\alpha+1}^{n}\left(n(1-\alpha)\right)^{-1} X_{(i)} - \sum_{i=1}^{n\alpha} b_i X_{(i)} - \sum_{i=n\alpha+1}^{n} b_i X_{(i)}
\end{aligned}$$

$$= \sum_{i=n\alpha+1}^{n} \left[\left(n(1-\alpha) \right)^{-1} - b_i \right] X_{(i)} - \sum_{i=1}^{n\alpha} b_i X_{(i)}$$

$$> \sum_{i=n\alpha+1}^{n} \left[\left(n(1-\alpha) \right)^{-1} - b_i \right] X_{(n\alpha)} - \sum_{i=1}^{n\alpha} b_i X_{(n\alpha)}$$

because in the first term all the $X_{(i)}$ are greater than $X_{(n\alpha)}$, and in the second term all the $X_{(i)}$ are less than or equal to $X_{(n\alpha)}$, so

$$\mathbf{c}'\mathbf{X}_{:\mathbf{n}} - \mathbf{c}'\mathbf{w}'\mathbf{X}_{:\mathbf{n}} > X_{(n\alpha)} \left\{ \sum_{i=n\alpha+1}^{n} \left[\left(n(1-\alpha) \right)^{-1} - b_i \right] - \sum_{i=1}^{n\alpha} b_i \right\} = 0.$$