Bias correction for estimated Distortion Risk Measure using the bootstrap

Joseph H.T. Kim*
University of Waterloo
200 University Ave. W.
Waterloo, Ontario, Canada N2L 3G1
May 25, 2009

Abstract

The bias of the empirical estimate of given risk measure has recently been of interest in the risk management literature. In particular, Kim and Hardy (2007) showed that the bias can be corrected for the Conditional Tail Expectation (CTE, a.k.a. Tail-VaR or Expected Shortfall) using the bootstrap. This article extends their result to the Distortion Risk Measure (DRM) class where the CTE is a special case. In particular, through the exact bootstrap, it is analytically proved that the bias of the empirical estimate of DRM with concave distortion function is negative and can be corrected using the bootstrap using the fact that the bootstrapped loss is majorized by the original loss vector. Since the class of DRM is a subset of the L-estimator class the result provides a sufficient condition for the bootstrap bias correction for L-estimators. Numerical examples are presented to show the effectiveness of the bootstrap bias correction. A practical guideline to choose the estimate with a lower mean squared error is also proposed, based on a single given sample, which would be useful in estimating risk measures.

Keywords: exact bootstrap; bias correction; distortion risk measure; majorization; conditional tail expectation.

*Email: jhtkim@uwaterloo.ca. The author acknowledges the support of the NSERC Canada.
1 Introduction

Risk measure, a mapping from a random variable to a non-negative real number:

\[ \rho(F) : X \rightarrow \mathcal{R}^+ , \]

where \( F \) is the underlying loss distribution, has traditionally been used by actuaries in premium setting. There are numerous candidates for \( \rho \): it could be the mean, often padded by dispersion measure such as the variance; one can use the quantile measure of the conditional tail risk measures. Recently its use has been extended to various places such as risk capital determination for banks and insurers, and international regulatory bodies now adopts policies that prescribe the level of minimum required capital based on a risk measure with a given safety level, such as Value-at-Risk at 99% or Conditional Tail Expectation at 95%\(^1\), where .

To estimate the given risk measure from internal model, actuaries typically rely on Monte Carlo simulations to produce loss data. The generated data are then used to estimate the risk measure, usually in the form of \( \rho(F_n) \), where \( F_n \) is the empirical distribution function. While it is evident that the empirical estimate \( \rho(F_n) \) becomes more accurate as \( n \) increases, practical constraints may prevent actuaries from obtaining more data points.

For large insurers the internal models can be quite complicated and slow to generate loss data. To hint the practical difficulty, consider a big multi-line insurer with millions of policies that tries to estimate the CTE at 95% based on its economic balance sheet that is market-consistent. For this insurer the market-consistent reserve of modern insurance product, calculated on a seriatim – policy by policy – basis, can create enough complexity depending on the contract options, covenants, and terms, sometimes leading to ‘nested’ stochastic simulations due to its market-consistent valuation. There are other numerous factors that make the simulation run time multiplied easily, such as, the interaction within asset classes that support contracts and between liabilities, projection period (can be one year, but could be several decades if projection is carried out until all liabilities have been settled), internal and external economic scenarios that may affect the value of assets, liabilities, and the dependence between them, different reporting system, and so on. Even though computers is getting ever fast, this situation is not likely to be resolved in the near future because the models become more sophisticated and complex due to the higher demand for more information. The time constraint of Monte Carlo simulation is well known (e.g., see Chueh (2002)) in the insurance community and often a major selling point addressed by commercial models. Also, there are situations where the number of data is impossible to increase. For example, datasets of floods or other natural disasters are by nature hard to obtain.

\(^{1}\)For a continuous loss distribution \( VaR_p(F) = F^{-1}(p) \), \( 0 < p < 1 \), the \( p \)-th quantile, and the Conditional Tail Expectation at confidence level \( p \) is defined by \( CTE_p(F) = E[X|X > F^{-1}(p)] \). Some authors call the Conditional Tail Expectation by different names: the Expected shortfall, the Tail-VaR, or Tail Conditional Expectation.
Under these circumstances actuaries are often expected to estimate risk measures from a sample of a relatively small size. As a response to this problem Kim and Hardy (2007) investigated the bias of the empirical estimates of the VaR and CTE. Using the exact bootstrap argument they showed that the bias of the empirical CTE is negative and, further, it can be corrected through the bootstrap. The same conclusion however does not hold for the VaR case, indicating that a blind application of the bootstrap in bias correction may fail. Therefore an assurance that the risk measure estimate actually is biased and the bootstrap can correct the bias is important in estimating risk measures, though it appears not easy to verify in general.

The bias of an estimated risk measure can be of interest itself but the variance would be the dominant component of its variability. Under certain contexts, however, importance of the bias becomes immense. For example, the current Canadian regulation on segregated fund contracts requires the CTEs of each policy, estimated from simulation, to be summed to produce the risk capital of the given portfolio (CIA Task Force, 2002). The bias of the estimated CTE thus accumulates – rather than canceling out – over the portfolio and, due to its negative bias, the resulting CTE estimate may lead to a substantial under-estimation of the total required capital. Thus the bias correction of Kim and Hardy (2007) has a direct application in this case. The bias of other risk measure estimates has also been discussed in Imü et al. (2005) and Centeno and Andrade e Silva (2005) in the literature.

The current article is a generalization of Kim and Hardy (2007). It investigates the bias of the estimated distortion risk measure (DRM) class, that contains the CTE as a special case. In particular, we show that the empirical DRM estimates, where the distortion function is concave, are negatively biased and its bias can be corrected through the bootstrap for a continuous loss distribution. The proof is analytic though the effectiveness of the bias correction varies over different risk measures, safety levels, as well as underlying models. It is also shown that the bias correction works for convex distortion as well, while the implication is limited in our context due to a lack of coherency. Since the DRM is equivalent to the L-estimator class, the result makes a contribution to the order statistics literature by providing a sufficient condition of bias correction using the bootstrap.

2 The Distortion Risk Measure

Among others the class of the distortion risk measures (DRM) defined for a distribution function $F(x)$ by

$$
\rho_g(F) = -\int_{-\infty}^{0} [1 - g(\bar{F}(x))]dx + \int_{0}^{+\infty} g(\bar{F}(x))dx
$$

(1)

has received much attention in the finance and actuarial literature; see Denuit et al. (2005) or Pflug and Römisch (2007) for a comprehensive treatment on risk measures. Here $g$, the
distortion function, is an increasing function defined on $[0, 1]$ with $g(0) = 0$ and $g(1) = 1$, and $\bar{F}(x) = 1 - F(x)$. It is assumed throughout the article that $F$ is a continuous loss distribution. The DRM is closely linked to the coherent axioms of risk measure by Artzner et al. (1999) for it satisfies all but one coherent axioms. If $g$ is concave the DRM further satisfies the subadditivity and becomes coherent; see, e.g., Wirch and Hardy (2000) and Dhaene et al. (2006). In the sequel we call the DRM with a concave distortion the concave DRM. By choosing a suitable $g$, one can easily express some popular risk measures through corresponding distortion function $g$. Some examples of continuous concave distortion functions corresponding to familiar risk measures are presented below:

Conditional Tail Expectation: $g^{CTE}(t) = \begin{cases} \frac{t}{1-\alpha} & \text{if } t \leq 1 - \alpha \\ 1 & \text{if } t > 1 - \alpha \end{cases}$

Proportional Hazard Transform: $g^{PHT}(t) = t^\beta$, $0 < \beta \leq 1$,

Dual-Power Transform: $g^{DPT}(t) = 1 - (1-t)^\beta$, $\beta \geq 1$,

and

Wang Transform: $g^{WT}(t) = \Phi(\Phi^{-1}(1) - \Phi^{-1}(\alpha))$,

where $\Phi(\cdot)$ is the distribution function of the standard normal. It is easy to verify that the coherency of the latter two measures by looking at the concavity of the corresponding distortion functions, while the distortion function of the VaR is a step function, failing to be concave.

Integrating (1) by parts leads to an alternative expression

\[
\rho_g(F) = \int_0^1 F^{-1}(1-u)dg(u) = \int_0^1 F^{-1}(u)g'(1-u)du,
\]

which clearly shows that the DRM belongs to the L-estimator class with the score function $g'(1 - \alpha)$ in the statistical literature. One natural candidate for the empirical estimate of (2) is obtained by replacing the true quantile with the sample quantiles to yield a linear combination of order statistics:

\[
\rho_g(F_n) = \sum_{i=1}^n \int_{(i-1)/n}^{i/n} X(i)g'(1-u)du = \sum_{i=1}^n X(i) \int_{(i-1)/n}^{i/n} g'(1-u)du
\]

\[
= \sum_{i=1}^n X(i) \left[ \left( \frac{n-i+1}{n} \right) - g \left( \frac{n-i}{n} \right) \right],
\]

where $X(i)$ is the $i$-th smallest order statistic from the sample $X_1, \ldots, X_n$. While this choice is a popular one in the literature, others are possible because sample quantiles are not uniquely defined.

In order to systematically investigate the bias of the empirical DRM estimate in a precise
manner for a given sample of size \( n \), we define the empirical distribution function of \( X \) based on a random sample \( X_1, \ldots, X_n \), as

\[
F_n(x) = \begin{cases} 
0, & \text{if } x < X_{(1)} \\
\frac{k}{n}, & \text{if } X_{(k)} \leq x < X_{(k+1)} \\
1, & \text{if } x \geq X_{(n)}.
\end{cases}
\]

It is known that this empirical distribution is unbiased for \( F \); see Rohatgi (1984). Consequently the empirical estimate of the DRM from its definition in (1), is given by

\[
\hat{\rho}_g = \rho_g(F_n) = -\int_{-\infty}^{0} [1 - g(F_n(x))]dx + \int_{0}^{+\infty} g(F_n(x))dx
\]

where \( F_n(x) = 1 - \hat{F}_n(x) \). To rewrite this in terms of the order statistics of a given sample, assume without loss of generality that \( X_{(k)} < T < X_{(k+1)} \). Then (4) becomes

\[
\hat{\rho}_g = -\int_{-\infty}^{X_{(1)}} [1 - g(\hat{F}_n(x))]dx - \sum_{i=2}^{k} \int_{X_{(i-1)}}^{X_{(i)}} [1 - g(\hat{F}_n(x))]dx - \int_{X_{(k)}}^{0} [1 - g(\hat{F}_n(x))]dx
\]

\[
+ \int_{0}^{X_{(k+1)}} g(\hat{F}_n(x))dx + \sum_{i=k+1}^{n-1} \int_{X_{(i)}}^{X_{(i+1)}} g(\hat{F}_n(x))dx + \int_{X_{(n)}}^{+\infty} g(\hat{F}_n(x))dx
\]

\[
= -\sum_{i=2}^{k} (X_{(i)} - X_{(i-1)}) \left[ 1 - g\left(1 - \frac{i}{n}\right)\right] - (0 - X_{(k)}) \left[ 1 - g\left(1 - \frac{k}{n}\right)\right]
\]

\[
+ (X_{(k+1)} - 0)g\left(1 - \frac{k}{n}\right) + \sum_{i=k+1}^{n-1} (X_{(i+1)} - X_{(i)})g\left(1 - \frac{i}{n}\right)
\]

\[
= \sum_{i=1}^{n} X_{(i)} \left[ g\left(\frac{n-i+1}{n}\right) - g\left(\frac{n-i}{n}\right)\right],
\]

where the last line uses \( g(0) = 0 \) and \( g(1) = 1 \). This shows that the empirical DRM estimate (4) coincides with the estimate given in (3).

Now we are ready to tackle the bias of the empirical DRM estimate. The following result is a straightforward generalization of the result of Centeno and Andrade e Silva (2005) where negative bias of the empirical estimate of the PHT was established.

**Proposition 2.1** The empirical estimate of the DRM defined in (4) is negatively biased for a concave distortion function \( g \).

**Proof:** First note that \( F_n(x) \) is an unbiased estimator of \( F(x) \), so that \( \mathcal{E}F_n(x) = F(x) \). Now the empirical estimate is given by

\[
\hat{\rho}_g = -\int_{-\infty}^{0} [1 - g(\hat{F}_n(x))]dx + \int_{0}^{+\infty} g(\hat{F}_n(x))dx.
\]
The expected value of the empirical estimate is then
\[ \mathcal{E}[\hat{\rho}_g] = \mathcal{E} \left[ - \int_{-\infty}^{0} [1 - g(\bar{F}_n(x))] dx + \int_{0}^{+\infty} g(\bar{F}_n(x)) dx \right] \]
\[ \leq - \int_{-\infty}^{0} [1 - g(\mathcal{E}\bar{F}_n(x))] dx + \int_{0}^{+\infty} g(\mathcal{E}\bar{F}_n(x)) dx \]
\[ = - \int_{-\infty}^{0} [1 - g(\bar{F}(x))] dx + \int_{0}^{+\infty} g(\bar{F}(x)) dx = \rho_g, \]
where the inequality comes from Jensen’s inequality with \( g \) being concave. The inequality is strict for strictly concave \( g \). \( \square \)

3 Bias correction using the bootstrap

To introduce the concept of the bootstrap suppose that we have an i.i.d sample of size \( n \) from an unknown distribution \( F \) and are interested in parameter \( \theta = \theta(F) \) whose empirical estimate is \( \hat{\theta} = \theta(F_n) \) with the empirical distribution function \( F_n \). The core idea of the nonparametric bootstrap is to repeatedly resample from the original sample, or the empirical distribution, with replacement. Let us now suppose that a bootstrap sample (or resample) of size \( n \) has been drawn from \( F_n \) with replacement. We denote this first bootstrap sample by \( F_n^{*}(1) \) where the superscript * indicates bootstrapping. Due to sampling with replacement some of the observations may be identical. After repeating this for \( R \) times, a series of bootstrap samples, \( F_n^{*}(1), F_n^{*}(2), \ldots, F_n^{*}(R) \), are obtained. Then the corresponding estimates \( \hat{\theta}_1^*, \ldots, \hat{\theta}_R^* \), where \( \hat{\theta}_i^* = \theta(F_n^{*}(i)) \), are produced to be used for various statistical inferences. Since \( F_n \) is treated as if it was the (unknown) population distribution function any inference can be possible to make, under regular assumptions. The bootstrap has successfully been applied to many statistical problems and become a widely-used non-parametric inference tool; see, e.g., Efron and Tibshirani (1993), Shao and Tu (1995), Hall (1992), and Davison and Hinkley (1997) for a comprehensive treatment for this subject; see these for the assumptions under which the bootstrap works. As \( n \) increases \( F_n \) approaches the true distribution \( F \) and bootstrap will be essentially the same as the Monte Carlo simulation. In theory \( \theta \) can be any parameter of interest including moments, quantiles, or functions of these.

As simple examples the bootstrap mean and variance estimates of \( \theta \), based on \( R \) resamplings, are given by
\[ \tilde{\theta}^* = R^{-1} \sum_{i=1}^{R} \hat{\theta}_i^* \]
and
\[ (R - 1)^{-1} \sum_{i=1}^{R} (\hat{\theta}_i^* - \tilde{\theta}^*)^2, \]
respectively. It is pedagogical to notice that these estimates are computed from a finite number of bootstrap simulations, thus are approximations of the true (or exact) bootstrap moment estimates, which can only be evaluated at $R = \infty$. If the evaluation at $R = \infty$ is somehow possible, we write the exact estimate of the bootstrap as $\mathcal{E}(\hat{\theta}^*|F_n) = \mathcal{E}(\theta(F_n^*)|F)$, which clearly indicates that the averaging is carried out over infinitely many different bootstrap samples obtained from the given empirical distribution $F_n$. In this case, of course, one has

$$\hat{\theta}^* = \mathcal{E}(\hat{\theta}^*|F_n),$$

as $R$ gets larger.

For this reason the bootstrap estimate $\hat{\theta}^*$ is generally subject to the resampling simulation error due to a finite size of $R$. The bootstrap has successfully been applied to many statistical problems and become a widely-used non-parametric inference tool; see, e.g., Efron and Tibshirani (1993), Shao and Tu (1995), Hall (1992), and Davison and Hinkley (1997) for a comprehensive treatment for this subject. The use of the bootstrap in risk measures can be found in, e.g., Kim and Hardy (2007), Kaiser and Brazauskas (2007), Kim and Hardy (2009), and Centeno and Andrade e Silva (2005).

It turns out that the exact bootstrap (EB), the bootstrap estimate evaluated at $R = \infty$, is available for the L-estimators, and the rest of this section is focused on developing some theoretical results on the bootstrap bias correction of the empirical DRM estimates in (5). To this extent, we first note the EB result of the L-estimator class by Hutson and Ernst (2000).

**Theorem 3.1 (Hutson and Ernst (2000))** The exact bootstrap (EB) of the estimate of $\mathcal{E}(X_{(r)}|F)$, $1 \leq r \leq n$ is

$$\mathcal{E}(X_{(r)}^*|F_n) = \sum_{j=1}^{n} w_{j(r)} X_{(j)},$$

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

$$B(x; a, b) = \int_{0}^{x} t^{a-1}(1-t)^{b-1}dt. \quad \Box$$

This result is easily extended to the L-estimator class, which is a linear combination of order statistics, and we notice from (5) that the empirical DRM estimate belongs to this class. For notational simplicity, we now define $X_n = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})^t$ and

$$c_i = g\left(\frac{n-i+1}{n}\right) - g\left(\frac{n-i}{n}\right)$$
to write the empirical DRM estimate in (5) as
\[ \hat{\rho}_g = \rho(F_n) = c_1X_{(1)} + c_2X_{(2)} + \ldots + c_nX_{(n)} = c'X_n, \tag{6} \]
where \( c = (c_1, c_2, \ldots, c_n)' \). Since \( \hat{\rho}_g \) is a linear combination of order statistics, its exact bootstrap (EB) estimate can be computed using the above theorem without resampling simulations. That is,
\[ \hat{\rho}_g^{EB} = \mathcal{E}(\hat{\rho}_g|F_n) = \mathcal{E}\left( \sum_{r=1}^{n} c_rX^*_{(r)}|F_n \right) = \sum_{r=1}^{n} c_r\mathcal{E}(X^*_{(r)}|F_n) = \sum_{r=1}^{n} \sum_{j=1}^{n} c_rw_j(r)X_{(j)}, \]
or more conveniently, in a matrix form,
\[ \hat{\rho}_g^{EB} = c'w'X_n, \]
where the elements of weight matrix \( w = \{w_j(r)\}_{j,r=1}^{n} \) are the EB weights for each element of \( X_n \). Consequently, the EB bias estimate and the bias-corrected estimate are, respectively,
\[ B = \mathcal{E}(\hat{\rho}_g^*|F_n) - \hat{\rho}_g = c'w'X_n - c'X_n, \tag{7} \]
and
\[ \hat{\rho}_g^{EB,bc} = \hat{\rho}_g - B = c'X_n - (c'w'X_n - c'X_n) = c'(2I - w')X_n. \tag{8} \]
For the CTE at \( \alpha \), a particular choice of \( c = (n(1-\alpha))^{-1}(0, \ldots, 0, 1, \ldots, 1)' \) with zeros for the first \( n\alpha \) elements, Kim and Hardy (2007) showed that the bias estimate in (7) is negative for any given sample.

In order to prove that the negative bias is true for all the concave DRM estimates, we now note that matrix \( w \) is doubly stochastic, meaning that all elements are nonnegative and the sum of each row and column is unity, as shown in Kim and Hardy (2007). Doubly stochastic matrices have been extensively utilized in many mathematical areas and are closely related to the concept of a particular partial order over vectors of real numbers, called majorization.

For two vectors \( a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n) \in \mathbb{R}^n \), we say \( b \) is majorized by \( a \) when
\[ \sum_{i=k}^{n} b_{(i)} \leq \sum_{i=k}^{n} a_{(i)}, \quad k = 1, 2, \ldots, n - 1, \]
\[ \sum_{i=1}^{n} b_{(i)} = \sum_{i=1}^{n} a_{(i)}, \]
where \( a_{(i)} \) is the \( i \)th smallest element of \( a \). The following classical result, e.g., from Marshall and Olkin (1979), is now in place for further developments.
Theorem 3.2 (Marshall and Olkin (1979)) An $n \times n$ matrix $P$ is doubly stochastic if and only if $Pa$ is majorized by $a$ for all $a \in \mathbb{R}^n$. □

Since $w$ is doubly stochastic, so is $w'$. Thus this result clearly shows that vector $w'X_n$ – which could be called the EB loss vector – is majorized by the original ordered loss vector $X_n$. Intuitively this indicates that the bootstrapped losses are weighted averages of the original losses where the weight is subscribed by $w$. Recognizing that averaging is a smoothing operation, this observation reveals important characteristics of the bootstrapping, such as being more robust than the empirical estimate for certain parameters. The following theorem is the main result of this section.

**Theorem 3.3** For a given loss sample, the empirical estimate of the concave DRM defined in (6) is bigger than its EB estimate. In other words, for any given sample or $F_n$,

$$
\mathcal{E}(\hat{\rho}^*_g|F_n) \leq \hat{\rho}_g
$$

for a concave distortion function $g$.

**Proof:** The assertion of the theorem can be rewritten as

$$
c'w'X_n \leq c'X_n.
$$

First we note that from the concavity of $g$,

$$
c_{i+1} - c_i = \left[ g\left(\frac{n-i}{n}\right) - g\left(\frac{n-i-1}{n}\right) \right] - \left[ g\left(\frac{n-i+1}{n}\right) - g\left(\frac{n-i}{n}\right) \right] = 2g\left(\frac{n-i}{n}\right) - \left[ g\left(\frac{n-i+1}{n}\right) + g\left(\frac{n-i-1}{n}\right) \right] \geq 0,
$$

implying that $0 \leq c_1 \leq c_2 \leq \ldots \leq c_n$.

Now write the EB loss vector as $X_n^* = w'X_n$, for notational simplicity in this proof. Then, from Theorem 3.2, we see that $X_n^*$ is majorized by $X_n$ because $w'$ is also doubly stochastic. In addition, the order is preserved in the sense that $X_{(1)}^* \leq \ldots \leq X_{(n)}^*$ because theses are bootstrapped quantiles. Therefore

$$
\sum_{i=k}^{n} X_{(i)}^* \leq \sum_{i=k}^{n} X_{(i)}, \quad k = 1, 2, \ldots, n-1,
$$

$$
\sum_{i=1}^{n} X_{(i)}^* = \sum_{i=1}^{n} X_{(i)}.
$$
Finally, by setting \( c_i = \sum_{j=1}^{i} (c_j - c_{j-1}) \) with \( c_0 = 0 \), we have

\[
c'X_n - c'w'X_n = c_1(X_{(1)} - X^*_{(1)}) + \ldots + c_1(X_{(n)} - X^*_{(n)})
\]

\[
= \sum_{i=1}^{n} c_i(X_{(i)} - X^*_{(i)})
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{i} (c_j - c_{j-1})(X_{(i)} - X^*_{(i)})
\]

\[
= \sum_{j=1}^{n} \sum_{i=j}^{n} (c_j - c_{j-1})(X_{(i)} - X^*_{(i)})
\]

\[
= \sum_{j=1}^{n} \left[ (c_j - c_{j-1}) \sum_{i=j}^{n} (X_{(i)} - X^*_{(i)}) \right] \geq 0,
\]

where the last inequality is justified by \( c_j \geq c_{j-1} \) and \( \sum_{i=j}^{n} (X_{(i)} - X^*_{(i)}) \geq 0 \), for \( j = 1 \), for each \( j = 1, \ldots, n \). \( \square \)

Since inequality

\[ \mathcal{E}(\hat{\rho}_g^*|F_n) \leq \hat{\rho}_g \]

holds for any given sample (that is, stochastically), it should hold on average too, to give

\[ \mathcal{E}(\mathcal{E}(\hat{\rho}_g^*|F_n)) \leq \mathcal{E}(\hat{\rho}_g), \]

or equivalently,

\[ \mathcal{E}(\hat{\rho}_g^{EB}) \leq \mathcal{E}(\hat{\rho}_g). \]

Consequently we establish the following DRM ’sandwich’ inequality for a concave distortion function \( g \):

\[ \mathcal{E}(\hat{\rho}_g^{EB}) \leq \mathcal{E}(\hat{\rho}_g) < \rho_g, \]

which generalizes the CTE bias result of Kim and Hardy (2007). The second inequality of the above formula comes from the proposition of the previous section. This proves therefore that the bootstrap bias correction for the coherent DRM works in the right direction because the unknown bias \( \mathcal{E}(\hat{\rho}_g) - \rho_g \) is estimated by the bootstrap bias \( \mathcal{E}(\hat{\rho}_g^*|F_n) - \hat{\rho}_g \).

In fact, the estimated bias (7), just like estimated risk measures, is itself biased and the standard bootstrap technique allows us to further adjust the bias using the double (or nested) bootstrap; see, e.g., Section 3.9 of Davison and Hinkley (1997) for details. To briefly explain this phenomenon consider the bias of \( \hat{\theta} = \theta(F_n) \), an arbitrary parameter estimate. Here the true bias

\[ \beta(F) = \mathcal{E}(\hat{\theta}) - \theta = \mathcal{E}(\theta(F_n)|F) - \theta(F) \]

is estimated by the bootstrap estimate

\[ B(F_n) = \mathcal{E}(\theta(F_n^*)|F_n) - \theta(F_n). \]
The above two expressions show a simple rule in the bootstrap plugin principle: The relationship between $F$ and $F_n$ is replaced with that of $F_n$ and $F^{**}_n$ in the bootstrap world. This is consistent with our intuition in that as $F_n$ is sampled from $F$, so is $F^{**}_n$ from $F_n$. Now if we look at $B(F_n)$ as an estimate of unknown parameter $\beta$, we see that it has its own bias, that is, $\mathcal{E}(B(F_n)|F) - \beta(F)$, and it can be estimated by the bootstrap:

$$C = \mathcal{E}(B(F^{**}_n)|F_n) - B(F_n).$$

Here $B(F^{**}_n)$ is given by, again using the same plugin principle, from (11),

$$B(F^{**}_n) = \mathcal{E}(\theta(F^{**}_n)|F^{**}_n) - \theta(F^{**}_n),$$

where $F^{**}_n$ is the empirical distribution of a sample drawn from $F^{**}_n$, that is, based on a bootstrap sample of the bootstrap sample. Using this so called the double bootstrap argument, we can express (12) as

$$C = \mathcal{E}(B(F^{**}_n)|F_n) - B(F_n)
\quad = \mathcal{E}[\mathcal{E}(\theta(F^{**}_n)|F^{**}_n) - \theta(F^{**}_n)|F_n] - [\mathcal{E}(\theta(F^{**}_n)|F_n) - \theta(F_n)]
\quad = \mathcal{E}[\mathcal{E}(\theta(F^{**}_n)|F^{**}_n)|F_n] - 2\mathcal{E}(\theta(F^{**}_n)|F_n) + \theta(F_n).$$

Consequently the bias-corrected estimate of $B(F_n)$, or adjusted estimate of the bias of $\theta(F_n)$, is given by

$$B^{adj} = B(F_n) - C
\quad = 3\mathcal{E}(\theta(F^{**}_n)|F_n) - 2\theta(F_n) - \mathcal{E}[\mathcal{E}(\theta(F^{**}_n)|F^{**}_n)|F_n].$$

In general the double bootstrapping is well-known for its heavy computation, but for the L-estimator class we can derive related quantities in a simple manner using the EB expression. For instance,

$$\mathcal{E}[\mathcal{E}(\hat{\rho}^{**}_n|F^{**}_n)|F_n] = E[c'w'X_{n}^*|F_n] = c'(w^2)'X_n,$$

so that the adjusted bias in (13) is shown to be

$$B^{adj} = B(F_n) - (c'(w^2)'X_n - 2c'w'X_n + c'X_n)
\quad = 3c'w'X_n - 2c'w'X_n - c'(w^2)'X_n.$$

In the next section we include the bias-corrected estimate using this adjusted bias $B^{adj}$ to see its performance. Before we investigate the numerical effectiveness of the bias correction under different loss models and risk measures, several remarks are in place:

**Remark 1:** When the distortion function $g$ is convex instead, then $c_1 \geq c_2 \geq \ldots \geq c_n$ and the inequalities in (9) reverse to lead to

$$\rho_g \leq \mathcal{E}(\hat{\rho}_g) \leq \mathcal{E}(\hat{\rho}_g^{EB}).$$
<table>
<thead>
<tr>
<th></th>
<th>Sample Size 200</th>
<th></th>
<th>Sample Size 800</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias (s.e.)</td>
<td>Std (s.e.)</td>
<td>rMSE</td>
<td></td>
</tr>
<tr>
<td>$\hat{CTE}$</td>
<td>-1.185(0.0986)</td>
<td>13.94(1.94)</td>
<td>13.99</td>
<td>0.2901(0.0499)</td>
</tr>
<tr>
<td>$\hat{CTE}^{OB}$</td>
<td>-2.48(0.097)</td>
<td>13.71(1.88)</td>
<td>13.936</td>
<td>-0.6109(0.0497)</td>
</tr>
<tr>
<td>$\hat{CTE}^{OB, bc}$</td>
<td>0.1087(0.101)</td>
<td>14.313(2.05)</td>
<td>14.313</td>
<td>0.03075(0.0503)</td>
</tr>
<tr>
<td>$\hat{CTE}^{EB}$</td>
<td>-2.479(0.0965)</td>
<td>13.649(1.86)</td>
<td>13.872</td>
<td>-0.6107(0.0497)</td>
</tr>
<tr>
<td>$\hat{CTE}^{EB, bc}$</td>
<td>0.1082(0.101)</td>
<td>14.244(2.03)</td>
<td>14.244</td>
<td>0.0305(0.0502)</td>
</tr>
<tr>
<td>$\hat{CTE}^{EB, bc2}$</td>
<td>0.09905(0.101)</td>
<td>14.254(2.03)</td>
<td>14.255</td>
<td>0.02972(0.0502)</td>
</tr>
</tbody>
</table>

Table 1: CTE 95% estimators for the GPD loss (True value: 99.8406)

**Remark 2:** It is often said that there is a tradeoff between bias and the variance when bootstrapped. While this phenomenon is often asserted and observed so in the bootstrap literature (e.g., Jeske and Sampath (2003) and Kim and Hardy (2007)) there has been no formal proof that the bias-correction necessarily increase the variance. For the L-estimator class we can take one step further in this direction using the above result. From Theorem 3.3 we now know that the EB estimate is stochastically smaller than the empirical estimate of the concave DRM. However the the comparison of variance requires a different stochastic ordering such as the dispersive ordering; see, e.g., Shaked and Shanthikumar (2007). For the concave DRM, one condition that warrants

$$\text{VAR}(\hat{\rho}_{g}^{EB}) \leq \text{VAR}(\hat{\rho}_{g}) \leq \text{VAR}(\hat{\rho}_{g}^{EB, bc})$$

is

$$\text{COV}(X_{(k+1)} - X_{(k)}, X_{(1)} + \ldots + X_{(i)}) \geq 0, \quad i = 1, \ldots, n, \quad k = 1, ..., n - 1,$$

which can essentially be derived from 3.H.4 of Marshall and Olkin (1979) based on the Schur-convexity of a real symmetric matrix. While this covariance inequality seems to be reasonable because the spacings between two consecutive order statistics and the partial sum of order statistics would move in a similar direction, little is known about the covariance of order statistics in general and it is unclear at this moment if (16) holds for all continuous distributions.

4 Numerical example

The first model for our numerical study are the Generalized Pareto Distribution (GPD), which has received much attention in recent insurance and finance applications. In particular the
<table>
<thead>
<tr>
<th></th>
<th>Sample Size 200</th>
<th></th>
<th>Sample Size 800</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias (s.e.)</td>
<td>Std (s.e.)</td>
<td>rMSE</td>
<td>Bias (s.e.)</td>
</tr>
<tr>
<td>( \hat{PHT} )</td>
<td>-0.1613(0.0157)</td>
<td>2.2205(0.0493)</td>
<td>2.2264</td>
<td>-0.05838(0.00793)</td>
</tr>
<tr>
<td>( \hat{PHT}^{OB} )</td>
<td>-0.2504(0.0156)</td>
<td>2.213(0.049)</td>
<td>2.2271</td>
<td>-0.09351(0.0079)</td>
</tr>
<tr>
<td>( \hat{PHT}^{OB.bc} )</td>
<td>-0.07226(0.0159)</td>
<td>2.2497(0.0506)</td>
<td>2.2509</td>
<td>-0.02326(0.00798)</td>
</tr>
<tr>
<td>( \hat{PHT}^{EB} )</td>
<td>-0.2511(0.0156)</td>
<td>2.2026(0.0485)</td>
<td>2.2169</td>
<td>-0.09367(0.0079)</td>
</tr>
<tr>
<td>( \hat{PHT}^{EB.bc} )</td>
<td>-0.07155(0.0158)</td>
<td>2.2391(0.0501)</td>
<td>2.2402</td>
<td>-0.0231(0.00797)</td>
</tr>
<tr>
<td>( \hat{PHT}^{EB.bc2} )</td>
<td>-0.05695(0.0159)</td>
<td>2.2436(0.0503)</td>
<td>2.2443</td>
<td>-0.01793(0.00798)</td>
</tr>
</tbody>
</table>

Table 2: \( \hat{PHT} \) at \( \beta = 0.85 \) estimators for the GPD loss (True value: 26.6667)

<table>
<thead>
<tr>
<th></th>
<th>Sample Size 200</th>
<th></th>
<th>Sample Size 800</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias (s.e.)</td>
<td>Std (s.e.)</td>
<td>rMSE</td>
<td>Bias (s.e.)</td>
</tr>
<tr>
<td>( \hat{DPT} )</td>
<td>-0.062(0.00958)</td>
<td>1.355(0.0184)</td>
<td>1.3564</td>
<td>-0.02076(0.00478)</td>
</tr>
<tr>
<td>( \hat{DPT}^{OB} )</td>
<td>-0.1112(0.00996)</td>
<td>1.3573(0.0184)</td>
<td>1.3618</td>
<td>-0.03331(0.00478)</td>
</tr>
<tr>
<td>( \hat{DPT}^{OB.bc} )</td>
<td>-0.01285(0.00966)</td>
<td>1.3661(0.0187)</td>
<td>1.3662</td>
<td>-0.008214(0.00478)</td>
</tr>
<tr>
<td>( \hat{DPT}^{EB} )</td>
<td>-0.1115(0.00955)</td>
<td>1.3507(0.0182)</td>
<td>1.3553</td>
<td>-0.03323(0.00477)</td>
</tr>
<tr>
<td>( \hat{DPT}^{EB.bc} )</td>
<td>-0.01247(0.00961)</td>
<td>1.3593(0.0185)</td>
<td>1.3593</td>
<td>-0.008298(0.00478)</td>
</tr>
<tr>
<td>( \hat{DPT}^{EB.bc2} )</td>
<td>-0.0121(0.00961)</td>
<td>1.3593(0.0185)</td>
<td>1.3594</td>
<td>-0.008274(0.00478)</td>
</tr>
</tbody>
</table>

Table 3: \( \hat{DPT} \) at \( \beta = 3 \) estimators for the Exponential loss (True value: 18.3333)

<table>
<thead>
<tr>
<th></th>
<th>Sample Size 200</th>
<th></th>
<th>Sample Size 800</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias (s.e.)</td>
<td>Std (s.e.)</td>
<td>rMSE</td>
<td>Bias (s.e.)</td>
</tr>
<tr>
<td>( \hat{PHT} )</td>
<td>-0.1992(0.00749)</td>
<td>1.059(0.0112)</td>
<td>1.0776</td>
<td>-0.08347(0.00382)</td>
</tr>
<tr>
<td>( \hat{PHT}^{OB} )</td>
<td>-0.2927(0.0074)</td>
<td>1.0472(0.011)</td>
<td>1.0874</td>
<td>-0.1221(0.00379)</td>
</tr>
<tr>
<td>( \hat{PHT}^{OB.bc} )</td>
<td>-0.1057(0.00765)</td>
<td>1.0819(0.0117)</td>
<td>1.0871</td>
<td>-0.04484(0.00386)</td>
</tr>
<tr>
<td>( \hat{PHT}^{EB} )</td>
<td>-0.2929(0.00738)</td>
<td>1.0435(0.0109)</td>
<td>1.0838</td>
<td>-0.1222(0.00379)</td>
</tr>
<tr>
<td>( \hat{PHT}^{EB.bc} )</td>
<td>-0.1056(0.00761)</td>
<td>1.0759(0.0116)</td>
<td>1.0811</td>
<td>-0.04468(0.00385)</td>
</tr>
<tr>
<td>( \hat{PHT}^{EB.bc2} )</td>
<td>-0.08916(0.00764)</td>
<td>1.0804(0.0117)</td>
<td>1.0841</td>
<td>-0.03846(0.00386)</td>
</tr>
</tbody>
</table>

Table 4: \( \hat{PHT} \) at \( \beta = 0.7 \) estimators for the Exponential loss (True value: 14.2857)
GPD is used in modeling the extreme observations as the limit distribution of excesses over high threshold. The distribution of GPD is defined as

\[ F_Y(y) = 1 - \left[ 1 + \frac{\xi y}{\sigma} \right]^{-1/\xi}, \quad \sigma > 0. \]

So the domain is \( y \geq 0 \) if \( \xi > 0 \), or \( 0 \leq y \leq -\sigma/\xi \) if \( \xi < 0 \); when \( \xi = 0 \) this reduces to the exponential distribution with mean \( \sigma \). Assuming \( \xi > 0 \), one can analytically derive some concave DRMs. For instance,

\[ CTE_\alpha(Y) = \frac{\sigma}{\xi} \left( \frac{(1 - \alpha)^{-\xi}}{1 - \xi} - 1 \right), \quad \xi < 1, \]

found in, e.g., Kim (2008), and

\[ PHT_\beta(Y) = \frac{\sigma}{\beta - \xi}, \quad 0 < \xi < \beta. \]

Our parameter choice is \( \sigma = 20 \) and \( \xi = 0.1 \) making the tail substantially heavy (only the first two moments exist with this choice of \( \xi \)). Our goal is to compare the performances of different estimates for two risk measures: the CTE at \( \alpha = 95\% \) and PHT at \( \beta = 0.85 \), in which case the true values are 99.8406 and 26.6667, from the above formulas, respectively. We added the CTE case even though it had already discussed in Kim and Hardy (2007) to examine the performance change due to double bootstrap.

The second model is an Exponential distribution with parameter \( \mu = 10 \), which coincides with the mean. For this relatively light-tailed loss model, we consider the DPT at \( \beta = 3 \) and PHT at \( \beta = 0.7 \), in which case the true values can be directly computed as 18.3333 and 14.2857, respectively.

The simulation study considers two sample sizes: \( n = 200 \) and 800, and for each sample size we generate 20,000 different sets of samples. For a single generated sample, we compute six different estimators for both risk measures: The empirical \( \hat{\rho} \), the EB estimate \( \hat{\rho}^{EB} \), and the bias-corrected EB estimate \( \hat{\rho}^{EB,bc} \), the bias-corrected EB estimate using the double bootstrap \( \hat{\rho}^{EB,bc2} \), the ordinary bootstrap (OB) estimate based on resampling simulation \( \hat{\rho}^{OB} \), and the bias-corrected bootstrap estimate based on resampling simulation \( \hat{\rho}^{OB,bc} \). The OB counterpart \( \hat{\rho}^{OB,bc2} \) is not considered for its computational burden. These estimates are averaged over 20,000 different samples to be compared on the mean squared error basis. The resampling simulation size for each generated sample is set at the half of the original sample size; that is, \( R = 100 \) for each sample of size 200 and \( R = 400 \) for each sample of size 800. The simulation result is given in Table 1 and 2 along with estimated standard errors. We have the following observations:

- For both risk measures where the distortion functions are concave, the empirical estimates are negatively biased and the EB estimates are further biased, confirming the DRM sandwich formula in (9).
• The bootstrap bias correction works for both risk measures, but at the expense of an increase in variance. This is a typical result reported in the literature; see

• The EB estimates are always better than the OB counterparts due to the fact that the resampling simulation error is completely eliminated.

• Further bias correction is achieved using the double bootstrap for both risk measures regardless of its sign. The bias-variance tradeoff however again applies here; that is, the bias correction using the double bootstrap further reduces bias but also increases variance from the first stage bootstrap bias-corrected estimate.

• In the first two tables, the EB estimate shows the lowest MSE, but the empirical is better in Table 3. This indicates that the optimal estimators among candidates may depend on models and the choice of risk measure.

5 A practical guideline based on a single sample

In practice the true risk measure is never known and thus the estimator with the smallest MSE cannot be selected. Kim and Hardy (2007) proposed a practical guideline to choose a better estimator between the empirical estimate and the EB one, using the bootstrap approximation for the unknown MSE. In this section we further generalize this guideline by replacing the unknown risk measure $\rho$ with the bias-corrected estimate, using the double bootstrap estimate $\hat{\rho}^{EB,bc2}$.

For any estimate of the DRM, say $\hat{\rho}$, its MSE is defined by

$$MSE(\rho^*) = (\mathcal{E}(\rho^*|F) - \rho)^2 + \mathcal{VAR}(\hat{\rho}|F),$$

where the true DRM $\rho$ and the distribution function $F$ are unknown. Now we approximate the MSE by plugging in the bootstrap estimates for the first two moments and replace $\rho$ with $\hat{\rho}^{EB,bc2}$, so that

$$\widehat{MSE}(\rho^*) = (\mathcal{E}(\rho^*|F_n) - \hat{\rho}^{EB,bc2})^2 + \mathcal{VAR}(\hat{\rho}|F_n).$$

This formula allows us to estimate the MSE from a single sample and thus leads to a better estimate in terms of the MSE. For the empirical and the EB estimates the approximation gives

$$\widehat{MSE}(\hat{\rho}) = (c'w'X_n - \hat{\rho}^{EB,bc2})^2 + c'\hat{\Sigma}_n c$$

and

$$\widehat{MSE}(\hat{\rho}^{EB}) = (c'(w^2)'X_n - \hat{\rho}^{EB,bc2})^2 + c'w'\hat{\Sigma}_n wc,$$

where $\hat{\Sigma}_n$ is the estimated covariance matrix of $\text{Cov}(X(i),X(j))$ based on the ordinary resamplings$^2$. To examine how this guideline works under different risk measures and models

$^2$Alternatively, for the variance components, one may use the non-parametric delta method, as discussed in Kim and Hardy (2009).
we use the identical simulated data sets as in the previous section, and applied the guideline for each sample to pick a better estimate between the empirical and the EB estimates. The results for 20,000 simulated samples of size 200 with resample size $R=600$ are shown in Table 5. The mixed estimate represents the performance of the estimate with smaller estimated $\nu$, produced by the guideline. The numbers show that the guideline works quite well in that its performance is ranked in the second place for the $\nu^\nu$ measure under the pym model, and ranked in the first place for the other cases. It is interesting to see that the mixed estimate sometimes performs even better than the both empirical and the EB estimates, advocating its advantage in minimizing the MSE; the guideline thus may be useful where the bias is non-cumulative across portfolio.
6 Concluding remarks

In risk measure estimation, usual Monte Carlo estimates based on relatively small sample size are known to be biased. In insurance applications, the bias at individual policy level could result in a substantial under- or over-estimation of portfolio level premium or capital determination. In this paper we consider the distortion risk measure (DRM) with concave distortion function, which encompasses popular risk measures including the CTE, Wang Transform, and PHT. In particular, we show that the bias of the empirical estimate of the concave DRM is negatively biased and can be reduced through the standard bootstrap bias-correction, at the cost of an increase in variance. A simulation study is carried out to illustrate the theoretical findings. Later a practical guideline to select a better estimate between the empirical and the EB estimates in the mean squared error criteria is also proposed based on the bootstrap approximation for a single sample. As an ancillary result, a sufficient condition that warrants the tradeoff between the bias correction and the variance increase is established based on the Schur-convexity of majorized vector.

References


