On the significance of expected shortfall as a coherent risk measure

Koji Inui a, Masaaki Kijima b,*

a Graduate School of Global Business, Meiji University, Chiyoda-ku, Tokyo 101-8301, Japan
b Graduate School of Economics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

Available online 29 September 2004

Abstract

This article shows that any coherent risk measure is given by a convex combination of expected shortfalls, and an expected shortfall (ES) is optimal in the sense that it gives the minimum value among the class of plausible coherent risk measures. Hence, it is of great practical interest to estimate the ES with given confidence level from the market data in a stable fashion. In this article, we propose an extrapolation method to estimate the ES of interest. Some numerical results are given to show the efficiency of our method.

Keywords: Value-at-Risk; Expected shortfall; Coherent risk; Historical simulation; Richardson’s extrapolation

1. Introduction

Value-at-Risk (VaR) is the most popular risk measure widely used by financial institutions all over the world. VaR measures the potential for significant loss in
a portfolio of financial assets, and VaR’s popularity is based on aggregation of several components of financial risk into a single number. However, VaR is often criticized, because it is not a coherent risk measure (to be defined later). That is, VaR does not in general satisfy the subadditivity condition. This means that diversification does not necessarily reduce VaR in contrast to the framework of modern portfolio theory. Also, VaR ignores statistical properties of the significant loss beyond the quantile point of interest.

Recently, practitioners start paying attention to the expected shortfall (hereafter ES) as an alternative risk measure. ES is a coherent risk measure, and calculates the conditional expected loss beyond VaR. ES was first proposed by Artzner et al. (1997), and further studied, among others, by Acerbi et al. (2001), Acerbi and Tasche (2002) and Acerbi (2002). Also, Rockafellar and Uryasev (2000) demonstrated some desired properties of ES when used for portfolio optimization. Ogryczak and Ruszczyński (2002) showed that the mean-risk model with ES as a risk measure is consistent with the second degree stochastic dominance relation.²

ES not only has these desired properties as a risk measure, but also plays a crucial role in the class of coherent risk measures. In particular, significance of ES as a risk measure is the following. Consider a real-valued random variable $X$ that represents the profit/loss of a portfolio over a given risk horizon. The cumulative distribution function (CDF for short) of $X$ is denoted by $F_X(x)$, and its inverse function by $F_X^{-1}(x)$, if it exists. Acerbi (2002) showed that any risk measure for $X$ represented as $\int_{F_X^{-1}(\lambda)}^1 F_X^{-1}(x) \phi(x) dx$ is coherent if and only if $\phi(x)$ is a probability density function (PDF for short) and decreasing in $x$. Using this result, it can be shown that any coherent risk measure can be represented as a convex combination of expected shortfalls with various confidence levels.

Turning to the estimation problem of a risk measure, historical simulation becomes popular for calculating VaR after the LTCM crisis. This is so, because the classical parametric method (the variance–covariance method) assumes that asset returns are normally distributed, but the LTCM crisis made clear that the normality cannot capture the fat-tailness of portfolio distribution and the non-linearity of portfolio VaR. The historical method uses the empirical distribution generated from the historical data, and VaR is estimated as a quantile point of the empirical (discrete) distribution. Hence, the historical method is free of model risk.³ Also, implementation of this method is very simple. Because of these reasons, many practitioners tend to use the historical simulation method with the help of bootstrap.

When the historical method is applied to ES, ES is estimated as an average of quantile points of the empirical distribution beyond VaR. Hence, the ES estimator may suffer from the same statistical drawback as the VaR estimator.⁴ Moreover,
Yamai and Yoshiba (2002) reported that the ES estimator is unstable especially for fat-tail distributions. Hence, it is of great practical interest to estimate ES with given confidence level in a stable manner using historical data from a profit/loss distribution.

This paper is organized as follows. In the next section, we provide the information necessary for what follows. Section 3 shows that any coherent risk measure can be represented as a convex combination of expected shortfalls with various confidence levels. Also, it is shown that an ES with given confidence level is optimal among a plausible class of coherent risk measures. In Section 4, we propose the Richardson extrapolation method for estimating the ES. Some numerical results are given to show the efficiency of our method. Section 5 concludes this paper.

2. Some preliminaries

Throughout this paper, we fix a probability space \((\Omega, \mathcal{F}, P)\) and consider a real-valued random variable \(X\) that represents the profit/loss of a portfolio over a given risk horizon. The CDF of \(X\) is denoted by \(F_X(x), x \in \mathbb{R}, \) i.e. \(F_X(x) = P\{X \leq x\},\) where \(\mathbb{R}\) stands for the real line. For the sake of simplicity of our presentation, it is assumed throughout that the CDF \(F_X(x)\) is absolutely continuous with PDF denoted by \(f_X(x) > 0\) for \(x \in \mathbb{R}\). This means that the CDF \(F_X(x)\) is strictly increasing in \(x \in \mathbb{R}\) and has the inverse function, which we denote by \(F_X^{-1}(x), \ 0 < x < 1.\)

We begin this section by the formal definition of VaR.

**Definition 1** (\(\alpha\)-quantile and VaR). For each \(\alpha, 0 < \alpha < 1,\) the \(\alpha\)-quantile of \(F_X(x)\) is given by

\[
F_X^{-1}(\alpha) = \inf \{x \mid F_X(x) \geq \alpha\}. \tag{1}
\]

VaR with 100(1 – \(\alpha\))% confidence level is defined by

\[
\text{VaR}_{(1-\alpha)} = -F_X^{-1}(\alpha) = -\inf \{x \mid F_X(x) \geq \alpha\}. \tag{2}
\]

From the definition, VaR is described as the possible maximum loss of a portfolio over a given risk horizon within a fixed confidence level. The VaR with 100(1 – \(\alpha\))% confidence level is nothing but the 100 \(\alpha\)-percentile of the profit/loss distribution \(F_X(x)\). Note the negative sign in Eq. (2); we need it, because the loss is usually a negative value. In practice, we consider VaR’s with confidence level of either 99% or 95%.

As pointed out earlier, VaR is often criticized, because it is not only a coherent risk measure (the definition is given in the next section) but also ignores statistical properties of the significant loss. Expected shortfall (ES) proposed by Artzner et al. (1997) is a coherent risk measure, and calculates the conditional mean loss beyond VaR. Many authors have studied the ES as an alternative risk measure. See, e.g., Acerbi et al. (2001), Acerbi and Tasche (2002) and Acerbi (2002).
Definition 2 (Expected shortfall). ES with 100(1 − α)% confidence level is defined by

\[
\text{ES}_{(1-\alpha)} = -\frac{1}{\alpha} \int_0^{x_{\alpha}} F_X^{-1}(p) \, dp.
\]

(3)

Let us denote the 100α-percentile of the profit/loss distribution \( F_X(x) \) by \( x_{\alpha} \), i.e. \( x_{\alpha} = -\text{VaR}_{(1-\alpha)} \). Then, it is readily seen that

\[
\text{ES}_{(1-\alpha)} = \text{VaR}_{(1-\alpha)} + \frac{1}{\alpha} \int_{-\infty}^{x_{\alpha}} F_X(x) \, dx = -\frac{1}{\alpha} \int_{-\infty}^{x_{\alpha}} x f_X(x) \, dx = -E[X \mid X \leq x_{\alpha}],
\]

since \( P\{X \leq x_{\alpha}\} = \alpha \) by definition. Hence, the ES calculates the conditional mean loss beyond VaR. Also, since \( \text{ES}_{(1-\alpha)} > \text{VaR}_{(1-\alpha)} \), ES is more conservative than VaR.

Now, suppose that we sample \( n \) observations from \( X \) independently. The samples are denoted by \( X_1, X_2, \ldots, X_n \). Note that they are IID (independent, identically distributed) samples with CDF \( F_X(x) \). Let us denote the corresponding order statistics by \( X_{(i)} : i = 1, 2, \ldots, n \), where \( X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)} \).

Definition 3 (Empirical distribution). The empirical distribution function generated from the samples \((X_1, X_2, \ldots, X_n)\) is defined by

\[
F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{x \geq X_i\}}, \quad x \in \mathbb{R},
\]

(4)

where \( \mathbf{1}_A \) denotes the indicator function of event \( A \) meaning that \( \mathbf{1}_A = 1 \) if \( A \) is true and \( \mathbf{1}_A = 0 \) otherwise.

Note that the empirical distribution function \( F_n(x) \) is not strictly increasing. Hence, its inverse function cannot be defined directly. In this paper, we consider lower (upper, respectively) empirical distribution value, which we denote by LEDV (UEDV), in the same manner as (1), whence the next definition.

Definition 4 (LEDV and UEDV). For given \( \alpha \) and samples \((X_1, X_2, \ldots, X_n)\), LEDV is defined by

\[
F_{\downarrow}^n(\alpha) = \inf \{ x \mid F_n(x) \geq \alpha \} = X_{k,n}, \quad \frac{k-1}{n} < \alpha \leq \frac{k}{n},
\]

(5)

while UEDV is defined by

\[
F_{\uparrow}^n(\alpha) = \sup \{ x \mid F_n(x) \geq \alpha \} = X_{k,n}, \quad \frac{k-1}{n} \leq \alpha < \frac{k}{n}.
\]

(6)

LEDV and UEDV are identical to \( X_{k,n} \) when \( n\alpha \) is not an integer, i.e. \( \frac{k-1}{n} < \alpha < \frac{k}{n} \). If \( n\alpha \) is an integer, then LEDV is equal to \( X_{n\alpha} \) while UEDV is equal to \( X_{n\alpha+1} \). Also, as \( F_n(x) \) is consistent to IID samples of the true CDF \( F_X(x) \), it is obvious that \( F_n(x) \) converges weakly to \( F_X(x) \) as \( n \to \infty \), hence both \( F_{\downarrow}^n(\alpha) \) and \( F_{\uparrow}^n(\alpha) \) converge weakly to \( F_{\downarrow}^{\downarrow}(\alpha) \) from below and above, respectively.

The VaR estimator is usually defined as LEDV, because it is more conservative than UEDV.
Definition 5 (VaR estimator). For given \( a \) and samples \((X_1, X_2, \ldots, X_n)\), let \( k \) be such that \((k - 1)h < a \leq kh\). Then, the VaR estimator with \(100(1 - a)\%\) confidence level is defined by

\[
\text{VaR}_{(1-a)} = -F^{-1}_n(a) = -X_{k:n},
\]  

where \(X_{k:n}\) denotes the \(k\)-th order statistics of the samples. In particular, if \(na\) is an integer, then the VaR estimator is given by \(\text{VaR}_{(1-a)} = -X_{na:n}\).

Suppose that \(na\) is an integer. Then, from (3), we have

\[
\text{ES}_{(1-a)} = -\frac{1}{a} \sum_{i=1}^{na} \int_{\frac{i}{na}}^{\frac{i+1}{na}} F^{-1}_X(p) \, dp.
\]

Hence, if we replace the (unknown) inverse function \(F^{-1}_X(p)\) by its empirical estimator \(F^{-1}_n(p) = X_{i:n}, \ (i - 1)h < p \leq ih,\) we obtain

\[
\text{ES}_{(1-a)} \approx -\frac{1}{na} \sum_{i=1}^{na} X_{i:n},
\]

An ES estimator is thus formally defined as follows. In the following, we denote the arithmetic mean of the order statistics by

\[
\overline{X}_{k:n} = \frac{X_{1:n} + X_{2:n} + \cdots + X_{k:n}}{k}, \quad k = 1, 2, \ldots, n.
\]

Definition 6 (ES estimator). Let \( k = [na] \), where \([x]\) denotes the largest integer not greater than \(x\), and let \( p = na - k \). ES estimator with \(100(1 - a)\%\) confidence level is defined by

\[
\text{ES}_{(1-a)} = \begin{cases} 
\overline{X}_{k:n}, & \text{if } na \text{ is an integer}, \\
-(1 - p)\overline{X}_{k:n} - p\overline{X}_{k+1:n}, & \text{if } na \text{ is not an integer}.
\end{cases}
\]

Remark 1. As Inui et al. (2003) pointed out, the VaR estimator (7) has a considerable positive bias, when the inverse function \(F^{-1}_X(p)\) is concave in \(p\) around small \(a\). On the other hand, a bias of the ES estimator (10) seems difficult to predict because of the rough discretization (8) of the inverse function \(F^{-1}_X(p)\) in \(p \in (0, a]\), although each quantile \(X_{i:n}\) has a negative bias (which contributes to the ES estimator as a positive bias). In Section 4, we use an extrapolation technique to adjust the bias and stabilize the ES estimator.

3. Characterization of coherent risk measures via ES

Artzner et al. (1999) present and justify the following four desirable properties for a measure of risk, \(\rho\) say:
1. **Monotonicity.** \( X, Y \in V, \ Y \geq X \Rightarrow \rho(Y) \leq \rho(X) \).
2. **Subadditivity.** \( X, Y, X + Y \in V \Rightarrow \rho(X + Y) \leq \rho(X) + \rho(Y) \).
3. **Positive homogeneity.** \( X \in V, \ h > 0, \ hX \in V \Rightarrow \rho(hX) = h\rho(X) \).
4. **Translation invariance.** \( X \in V, \ a \in \mathbb{R} \Rightarrow \rho(X + a) = \rho(X) + a \).

A risk measure that satisfies these conditions is called a coherent risk measure. Here, \( V \) stands for a class of random variables under consideration.

Let \( \mathcal{A} \) be a class of functions \( \Phi(p) \) defined on \([0,1]\) such that \( \Phi(p) \geq 0 \), \( \int_0^1 \Phi(p) \, dp = 1 \) and \( \Phi(p) \) is decreasing in \( p \). For the profit/loss of a portfolio with CDF \( F_X(x) \), define the spectral risk measure of \( X \):

\[
M_X = -\int_0^1 F_X^{-1}(p)\Phi(p) \, dp,
\]

for some \( \Phi \in \mathcal{A} \). The spectral risk measure \( M_X \) is first considered by Acerbi (2002). Of significance in spectral risk measures is that \( M_X \) is coherent if and only if \( \Phi \in \mathcal{A} \).

Now, from (3) and integration by parts, it is readily seen that

\[
-\int_0^1 F_X^{-1}(p)\Phi(p) \, dp = \int_0^1 ES_{1-x} \, d\mu(x) - \Phi(1)E[X],
\]

where \( d\mu(x) = -x \, d\Phi(x) \). Note that, since \( \Phi \in \mathcal{A} \), \( \Phi(1) = 1 \) implies \( \Phi(p) = 1 \) for all \( p \in [0,1] \). In the following, we assume for simplicity of our presentation that \( \Phi(1) = 0 \). Then, \( d\mu(x) \geq 0 \) and \( \int_0^1 d\mu(x) = 1 \). Hence, \( \mu(x) \) can be considered to be a CDF of a random variable defined on \([0,1]\). Let \( Y \) be such a random variable. Then, we have the following characterization of a coherent risk measure.

**Proposition 1.** Suppose \( \Phi(1) = 0 \) in the representation (11). A risk measure \( M_X \) for \( X \) is coherent if and only if it is represented as

\[
M_X = E[ES_{1-Y}]
\]

for some random variable \( Y \), independent of \( X \) and defined on \([0,1]\). Hence, any coherent spectral risk measure is given by a convex combination of expected shortfalls.

Let \( g(x) < 0, \ 0 < x < x^* \) for some \( x^* \), and suppose that \( g(x) \) is increasing, concave in \( x \). Define \( \eta(x) = \int_0^x g(p) \, dp / x \). Then

\[
\eta'(x) = \frac{xg(x) - \int_0^x g(p) \, dp}{x^2} > 0,
\]

since \( g(x) \) is increasing and negative on \((0, x^*)\). Similarly, we obtain

\[
\eta''(x) = \frac{x^2g'(x) - 2xg(x) + 2\int_0^x g(p) \, dp}{x^3}.
\]

Note that

\[
xg(x) - \int_0^x g(p) \, dp = \int_0^x (g(x) - g(p)) \, dp = \int_0^x \left( \int_p^x g'(y) \, dy \right) \, dp,
\]
from which we obtain

\[ xg(x) - \int_0^x g(p) \, dp = \int_0^x yg'(y) \, dy, \]

after the change of the order of integrations. It follows that

\[ \eta''(x) = 2 \int_0^x y(g'(x) - g'(y)) \, dy, \]

which is non-positive, because \( g(x) \) is concave. Therefore, under the given conditions, the function \( \eta(x) \) is increasing, concave in \( x \in (0,x^*) \).

Suppose that \( F_X^{-1}(x) \) is increasing, concave in \( x \in (0,x^*) \) for some \( x^* < 1 \). Then, by definition (3), \( ES_{(1-x)} \) is increasing and convex for sufficiently small \( x \). Let \( \mathcal{B}_x = \{ Y : E[Y] = x, 0 < Y < x^* \} \). By Jensen’s inequality,

\[ M_X = E[ES_{(1-Y)}] \geq ES_{(1-x)} \quad Y \in \mathcal{B}_x, \]

provided that \( F_X^{-1}(x) \) is increasing, concave in \( x \in (0,x^*) \). We thus have the following.

**Proposition 2.** Suppose that \( F_X^{-1}(x) \) is increasing, concave in \( x \in (0,x^*) \). Then, \( ES_{(1-x)} \) is optimal among \( Y \in \mathcal{B}_x \) in the sense that

\[ ES_{(1-x)} = \inf_{Y \in \mathcal{B}_x} E[ES_{(1-Y)}]. \]

It should be noted that, if the CDF \( F_X(x) \) is convex in \((\infty,X^*)\) for some \( X^* \), \( F_X^{-1}(x) \) is concave in \( x \in (0,x^*) \) with \( x^* = F_X^{-1}(X^*) \). Many CDF’s are convex in \((\infty,X^*)\) for small \( X^* \).

For some integer \( n \), we have from (11) that

\[ M_X = -\sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} F_X^{-1}(p) \phi(p) \, dp. \]

If the (unknown) inverse function \( F_X^{-1}(p) \) is replaced by its empirical estimator \( F^{-1}_n(p) = X_{i:n}, (i-1)/n < p \leq i/n \) we obtain

\[ M_X \approx -\sum_{i=1}^n X_{i:n} \phi_i, \quad \phi_i = \int_{\frac{i-1}{n}}^{\frac{i}{n}} \phi(p) \, dp. \] \hspace{1cm} (12)

Note that, since \( \phi(p) \geq 0, \int_0^1 \phi(p) \, dp = 1 \) and \( \phi(p) \) is decreasing in \( p \), the sequence \( \{ \phi_i \} \) also satisfies the same properties. Namely, \( \phi_1 \geq 0, \sum_{i=1}^n \phi_i = 1 \) and \( \phi_i \) is decreasing in \( i \).

Let \( \mathcal{A}^d \) be a class of sequences \( \{ \phi_i \} \) such that \( \phi_i \geq 0, \sum_{i=1}^n \phi_i = 1 \) is decreasing in \( i \). For \( \{ \phi_i \} \in \mathcal{A}^d \), define

\[ M^d_X = -\sum_{i=1}^n X_{i:n} \phi_i, \quad \{ \phi_i \} \in \mathcal{A}^d. \] \hspace{1cm} (13)

The risk measure is called the discrete spectral risk measure for \( X \). The ES estimator given by (10) is the discrete spectral risk measure with \( \phi_i = 1/k \) for \( i = 1,\ldots,k \) and \( \phi_i = p/k \) for \( i = k+1 \), where \( k = [nx] \), if \( nx \) is not an integer.
Remark 2. A convex combination of order statistics $X_{i:n}$ is called an L-estimator. That is, an L-estimator is given by $\sum_{i=1}^{n} w_i X_{i:n}$ for some $w_i \geq 0$ with $\sum_{i=1}^{n} w_i = 1$. Hence, the discrete spectral risk measure $M^d_X$ is an L-estimator with the weights $\phi_i$ being decreasing in $i$.

4. Estimation of ES via extrapolation

In the previous section, we observed that any coherent spectral risk measure is characterized as a convex combination of expected shortfalls (ES’s). Also, ES is optimal in the sense that it gives the minimum value among the class of plausible coherent risk measures. Hence, it is of great practical interest to estimate the ES with given confidence level from the market data in a stable fashion. In this section, we propose an extrapolation method to estimate the ES of interest. Some numerical results are also given to show the efficiency of our method.

An extrapolation is the transformation method that transforms the original sequence $\{s_n\}$ to another $\{t_n\}$ with the same limit which converges faster than the original one. In the finance literature, Geske and Johnson (1984) applied Richardson’s extrapolation to price American put options, and it is known that this method provides a sufficiently accurate approximation.

4.1. Richardson’s extrapolation method

In this subsection, we explain Richardson’s extrapolation method for the completeness of our presentation. Let $\{s_n\}$ be a sequence that converges to $s$, and consider $(k + 1)$-dimensional simultaneous equations (with respect to $t$ and $a_i$) and a properly selected sequence $\{x_n\}$ such that

$$s_{n+i} = t + a_1 x_{n+i} + a_2 x^2_{n+i} + \cdots + a_k x^k_{n+i}, \quad i = 0, 1, \ldots, k. \tag{14}$$

From Cramer’s formula, we obtain

$$t = t^{(k)}_n = \begin{vmatrix} s_n & s_{n+1} & \cdots & s_{n+k} \\ x_n & x_{n+1} & \cdots & x_{n+k} \\ \vdots & \vdots & \ddots & \vdots \\ x^k_n & x^k_{n+1} & \cdots & x^k_{n+k} \\ 1 & 1 & \cdots & 1 \end{vmatrix}, \quad n = 1, 2, \ldots. \tag{15}$$
This means that we construct new sequences \( \{t_n^{(k)}\} \) with \( \{t_n^{(0)}\} = \{s_n\} \) in which \( \{t_n^{(k+1)}\} \) converges to \( s \) faster than \( \{t_n^{(k)}\} \) when the sequence \( \{x_n\} \) satisfies regularity conditions. \(^5\)

Suppose that \( x_n = 1/n \) with \( k = 2 \) in (15). Then, for example,

\[
t_1^{(2)} = \begin{bmatrix} s_1 & s_2 & s_3 \\ 1 & 1/2 & 1/3 \\ 1 & 1/4 & 1/9 \\ 1 & 1 & 1 \\ 1 & 1/2 & 1/3 \\ 1 & 1/4 & 1/9 \end{bmatrix} = \frac{1}{2} s_1 - 4s_2 + \frac{9}{2} s_3.
\]

(16)

A higher order of \( \{t_n^{(k)}\} \) can be similarly obtained.

Now, suppose that we obtain a sequence of ES estimators \( s_n = ES_{1-x}^{(a)} \) that converges to the true ES value \( ES_{1-x}^{(a)} \). According to Richardson’s extrapolation method, the sequence \( \{t_n^{(2)}\} \) calculated by (16) converges to \( ES_{1-x}^{(a)} \) faster than \( s_n \). Hence, our approximation for the true ES value will be \( t_n^{(2)} \) for some \( n \). In our numerical experiments, we choose \( n = 1 \).

### 4.2. Simulation procedure

In this numerical experiment, we use \( t \)-distributions with degree of freedom (DF for short) \( \text{DF} = 3, 4 \) and 5. The smaller the DF, the fatter the tail of the distribution. Empirical distributions are generated by simple Monte Carlo simulation. As a comparison, the case of normal distributions is also examined. Throughout this simulation, we fix the confidence level \( x = 0.01 \).

Formally, our simulation procedure is as follows:

1. Generate the empirical distribution with 300 samples.
2. Resample \( N \) different (non-duplicated) samples from the original dataset, if necessary, where \( N = 100, 200 \) and 300, and calculate the ES by (10) for each case. \(^6\)
3. Repeat Step 2 for 1,000 times to calculate the mean of ES’s, which is used as \( s_n \), where \( n = 1 \) (2 and 3, respectively) corresponds to the case of \( N = 100 \) (200 and 300).
4. Apply Richardson’s extrapolation (16) to obtain \( t_1^{(2)} \), which is used as an estimator \( \hat{ES}_{0.99} \) for the value \( ES_{0.99}^{(a)} \).
5. Repeat Step 1 to Step 4 for 1000 times and calculate the basic statistics of the estimator \( \hat{ES}_{0.99} \).

These results are compared with the regular estimator of ES, defined as \( \overline{X}_{k,n} = s_4 \) in the definition (10).

\(^5\) For example, it is sufficient that \( 1 > x_1 > x_2 > \cdots \) and \( \lim_{n \to \infty} x_n = 0 \). See Brezinski and Redivo Zaglia (1991) for details.

\(^6\) Of course, we do not need to resample for the case \( N = 300 \).
4.3. Simulation results

Table 1 presents the basic statistics of the ES estimators with confidence level 99% \((\alpha = 0.01)\), where ES stands for the ordinary ES estimator calculated by (10) and \(\hat{\text{ES}}_1\), the extrapolation estimator calculated by (16). As comparison, the theoretical ES, denoted by \(\text{ES}^*\), is also appended. The bias is calculated by

\[
\text{Bias} = \frac{\text{Mean of ES} - \text{ES}^*}{\text{ES}^*}.
\]

The bigger the number, the more biased the estimator is. The column ‘MSE’ stands for the mean squared error. As an alternative extrapolation estimator, we also calculate the case with \(k = 1\) and \(n = 2\) in (15). The obtained estimation is denoted by \(\hat{\text{ES}}_2\). The performance of extrapolation can be understood by a comparison with the ordinary ES estimator. From this table, the extrapolation method works very well to adjust the bias for every case.

4.4. The case of VaR

The extrapolation idea can be applied to VaR estimators as well. Table 2 shows the same statistics of the VaR estimators with confidence level 99% \((\alpha = 0.01)\) as of the ES estimators. Here, we test the two types of VaR estimators given in

7 The theoretical values are calculated by numerical integration. Hence, these values may be subject to numerical error.
Definition 4, the one being applied to LEDV, denoted by $\widehat{\text{VaR}}^L$, and the other to UEDV, denoted by $\widehat{\text{VaR}}^U$. In Table 2, it is explicitly observed that the ordinary VaR, denoted by VaR in the table, have positive bias, as proved in Inui et al. (2003). The bias increases as the DF of the underlying $t$-distribution decrease (the degree of fat-tailness increase).

The extrapolation method works well even for the VaR estimators. In particular, $\widehat{\text{VaR}}^U$ provides better results. Note that VaR is usually estimated by LEDV. However, when the extrapolation method is applied, we should use UEDV in order to obtain accurate and stable VaR estimates.

5. Conclusion

In this paper, we show that any coherent spectral risk measure is given by a convex combination of expected shortfalls, and an expected shortfall is optimal in the sense that it gives the minimum value among the class of plausible coherent risk measures. Hence, it is of great practical interest to estimate the ES with given confidence level from the market data in a stable fashion. This paper proposes an extrapolation method to estimate the ES of interest.

The extrapolation method seems very effective even for cases with strong bias. In particular, strong bias for VaR estimators can be reduced considerably by the extrapolation method, if we use UECV as possible quantile estimators. In the case of ES estimators, the improvement is moderate, although the remaining bias is only less than 1%.

Table 2
VaR estimates by 300 samples with 99% confidence level

<table>
<thead>
<tr>
<th></th>
<th>$\text{VaR}^*$</th>
<th>$\text{VaR}$</th>
<th>$\widehat{\text{VaR}}^L_{(1)}$</th>
<th>$\widehat{\text{VaR}}^L_{(2)}$</th>
<th>$\widehat{\text{VaR}}^U_{(1)}$</th>
<th>$\widehat{\text{VaR}}^U_{(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>2.326</td>
<td>2.381</td>
<td>2.301</td>
<td>2.308</td>
<td>2.321</td>
<td>2.316</td>
</tr>
<tr>
<td></td>
<td>0.233</td>
<td>0.421</td>
<td>0.321</td>
<td>0.319</td>
<td>0.267</td>
<td></td>
</tr>
<tr>
<td>Bias (%)</td>
<td>2.06</td>
<td>-0.96</td>
<td>-0.68</td>
<td>-0.22</td>
<td>-0.41</td>
<td></td>
</tr>
<tr>
<td>$t(DF = 3)$</td>
<td>4.541</td>
<td>4.950</td>
<td>4.618</td>
<td>4.456</td>
<td>4.524</td>
<td>4.510</td>
</tr>
<tr>
<td></td>
<td>1.331</td>
<td>2.677</td>
<td>1.884</td>
<td>1.629</td>
<td>1.327</td>
<td></td>
</tr>
<tr>
<td>Bias (%)</td>
<td>5.85</td>
<td>1.11</td>
<td>-1.20</td>
<td>-0.24</td>
<td>-0.44</td>
<td></td>
</tr>
<tr>
<td>$t(DF = 4)$</td>
<td>3.747</td>
<td>4.017</td>
<td>3.762</td>
<td>3.705</td>
<td>3.759</td>
<td>3.748</td>
</tr>
<tr>
<td></td>
<td>0.846</td>
<td>1.590</td>
<td>1.173</td>
<td>1.028</td>
<td>0.842</td>
<td></td>
</tr>
<tr>
<td>Bias (%)</td>
<td>5.18</td>
<td>0.29</td>
<td>-0.80</td>
<td>0.23</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.650</td>
<td>1.190</td>
<td>0.868</td>
<td>0.829</td>
<td>0.684</td>
<td></td>
</tr>
<tr>
<td>Bias (%)</td>
<td>4.70</td>
<td>-0.79</td>
<td>-0.95</td>
<td>1.04</td>
<td>0.59</td>
<td></td>
</tr>
</tbody>
</table>
References