Nonparametric Estimation of Conditional Expected Shortfall Admitting a Location-Scale Model

Emmanuel Torsen

Department of Mathematics, Pan African University Institute of Basic Sciences, Technology, and Innovation, Kenya
&
Department of Statistics and Operations Research
Modibbo Adama University of Technology Yola, Nigeria

Peter N. Mwita

Department of Mathematics, Machakos University, Kenya

Joseph K. Mung’atu

Department of Statistics and Actuarial Sciences
Jomo Kenyatta University of Agriculture and Technology, Kenya

This article is distributed under the Creative Commons by-nc-nd Attribution License. Copyright © 2019 Hikari Ltd.

Abstract

The Expected Shortfall as the most preferred measure of risk takes into account all the possible losses that exceed the severity level corresponding to the Value-at-Risk. In this paper at hand, we have proposed a nonparametric method of estimation for conditional expected shortfall assuming that the returns on an asset or portfolio admit a location-scale model. We have verified the properties of the estimator through simulation using the bootstrap method, the variance, bias, and asymptotic mean square error are relatively small, meaning that the estimator performs relatively good. Financial market data for TOTAL company quoted on the Nigerian Stock Exchange was used to illustrate the applicability of the estimator.

Keywords: Expected Shortfall, Location-Scale Model, Nonparametric Estimation, Three-Step, Conditional Value-at-Risk
1 Introduction

A major concern for financial institutions owners and regulators is the risk analysis. Value-at-Risk (VaR) is one of the most used and common measures of risk used in finance [6]. It measures the down-side risk and is determined for a given probability level $\tau$. Typically, in measuring losses, VaR is the lowest value which exceeds this level (the quantile of the loss distributions). The Expected Shortfall (ES) due to [1] is the average of the $100(1-\tau)$% worst losses. The ES takes into account all the possible losses that exceeds the severity level corresponding to the VaR. Since the first Basel Accord [15], the VaR and recently, the ES [3] forms the essential basis for the determination of market risk capital [12]. [2] proposed ES to solve the problems inherent in VaR. ES is the conditional expectation of loss given that the loss is beyond the VaR level, as defined in section (1.2). The ES shows the average loss when the loss exceeds the VaR level. In other words, ES is a numerical figure indicating the average of the possible losses that could be incurred beyond a selected tolerance level, a vast literature refer to this as the Tail Conditional Expectation (TCE), Conditional Loss (CL) or Tail Loss (TL). Taking the ES implies a selection of a benchmark value (VaR). The values beyond this benchmark value (VaR) are then averaged to come up with the ES [16].

ES and CES have been known and popular for more than two decades now among actuary sciences and industries, see [13, 4]. There exists a large studies in the literature on nonparametric estimation of unconditional VaR and ES, for example., [14], [5], [4], [10] and [9].

2 Estimation of Conditional Expected Shortfall (CES)

Here, the estimation of $CES(X)\tau$ for processes $Y_t$ that admit a location-scale representation given as

$$Y_t = m(X_t) + \sqrt{h(X_t)}\epsilon_t \quad (1)$$

where $m$ and $h > 0$ are nonparametric functions defined on the range of $X_t$, $\epsilon_t$ is independent of $X_t$, and $\epsilon_t$ is an independent and identically distributed (iid) innovation process with $\mathbb{E}(\epsilon_t) = 0$, $\text{Var}(\epsilon_t) = 1$ and the unknown distribution function $F_{\epsilon}$.

From equation (1) we have;

$$CVaR(X)\tau := Q_{Y|X}(\tau|x) = m(X_t) + \sqrt{h(X_t)}q(\tau) \quad (2)$$

where $Q_{Y|X}(\tau|x)$ is the conditional $\tau-$quantile associated with $F(y|x)$ and $q(\tau)$ is the $\tau-$quantile associated with the error innovation $F_{\epsilon}$. The estimator
of (2), its properties, and the prediction intervals has been studied in [18] and [20] respectively. Hence,

\[
CES(X)_\tau \equiv \mathbb{E}(Y_t/Y_t > Q_{Y|X}(\tau|x), \bar{X}_t = x) = m(X_t) + \sqrt{h(X_t)} \mathbb{E}(\varepsilon_t|\varepsilon_t > q(\tau))
\tag{3}
\]

Where \(Q_{Y|X}(\tau|x)\) is the conditional \(\tau\)–quantile associated with \(F_{Y|X}(y|x)\) and \(q(\tau)\) is the \(\tau\)–quantile associated with \(F_{\varepsilon}\).

The problem of estimating \(m(X)\) and \(h(X)\) in equation (3) was studied by [7] and [8]. Now, with the estimators of the mean and variance functions, we have the sequence of squared residuals \(\{r_i = (Y_i - \hat{m}(x))^2\}_{i=1}^n\).

These estimators the mean and variance functions are then used to get a sequence of Standardized Nonparametric Residuals (SNR) \(\{\hat{\varepsilon}_i\}_{i=1}^n\), where;

\[
\hat{\varepsilon}_i = \begin{cases} 
\frac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}(X)}}, & \text{if } \hat{h}(X) > 0 \\
0, & \text{if } \hat{h}(X) \leq 0
\end{cases}
\tag{4}
\]

We use these SNR to obtain the cumulative conditional density estimator of \(F_{\varepsilon}\), see our previous paper for details [19].

With the estimators of the mean function \(m(X)\), the variance function \(h(X)\) and the unknown error innovation, our estimator for Conditional Value-at-Risk (CVaR), discussed in [18] is given as;

\[
\bar{CVaR}(X)_\tau := \hat{Q}_{Y|X}(\tau|x) = \hat{m}(x) + \hat{h}^{1/2}(x)\hat{q}(\tau)
\tag{5}
\]

and hence, our estimator for Conditional Expected Shortfall is

\[
\bar{CES}(X)_\tau := \mathbb{E}(Y_t/Y_t > \hat{Q}_{Y|X}(\tau|x), \bar{X}_t = x) = \hat{m}(x) + \hat{h}^{1/2}(x)\mathbb{E}(\varepsilon_t|\varepsilon_t > \hat{q}(\tau))
\tag{6}
\]

To discussed the asymptotic properties of (6), the estimator of (3), to do this we make the following basic assumptions:

### 2.1 Assumptions

**A: Bandwidth**

1. \(b \to 0\), as \(n \to \infty\)

2. \(nb \to \infty\), as \(n \to \infty\)

**B: Kernel**
1. $K$ has compact support
2. $K$ is symmetric
3. $K$ is Lipschitz continuous
4. $K$ is $\int_{-\infty}^{\infty} K(u)du = 1$ and $\int_{-\infty}^{\infty} uK(u)du = 0$ with $\mu_2(K) = \int_{-\infty}^{\infty} u^2K(u)du$ and $R(K) = \int_{-\infty}^{\infty} K(u)^2du$ being the second moment (Variance) and Roughness of the kernel function respectively.
5. $K$ is bounded and there is $\bar{K} \in \mathbb{R}$, with $K(u) \leq \bar{K} < \infty$ and $K(u) \geq 0, \forall u \in \mathbb{R}$

3 Asymptotic Properties of the Estimator

To examine the asymptotic properties of our estimator, we quickly mention two results of importance from [7, 8], see also [19] and [18]

In Theorem 1 of [8], under the assumptions of the aforementioned paper,

$$\sqrt{nb} \left[ \hat{h}(x) - h(x) - \text{Bias}(\hat{h}(x)) \right] \overset{d}{\to} \mathcal{N}(0, f^{-1}(x)h^2(x)\lambda^2(x)\int k^2(u)du)$$

where

$$\lambda^2(x) = \mathbb{E}[(\epsilon^2 - 1)^2|X = x], \quad \epsilon = \frac{Y-m(X)}{h(X)}$$

$$\mu_2(k) = \int u^2K(u)du$$

$\text{Bias}(\hat{h}(x)) = \frac{b^2}{2} \mu_2(k)h''(x)^2$

This means that

$$\hat{h}(x) \overset{d}{\to} h(x)$$

with

$$\mathbb{E}(\hat{h}(x)) = h(x) + \frac{b^2}{2} \mu_2(k)h''(x)^2 = M_h$$

$$\text{Var}(\hat{h}(x)) = \frac{1}{nb^2(x)}R(k)h^2(x)\lambda^2(x) = V_h$$

Hence

$$\hat{h}(x) \sim \mathcal{N}(M_h, V_h)$$

Also in [7], we have that

$$\hat{m}(x) \overset{d}{\to} \mathcal{N}(M_{\hat{m}}, V_{\hat{m}})$$

where

$$M_{\hat{m}} = m(x) + \frac{b^2}{2} m''(x)\mu_2(k), \quad V_{\hat{m}} = \frac{\sigma^2(x)}{nb^2(x)} \int_{-\infty}^{\infty} K^2(u)du = \frac{\sigma^2(x)R(k)}{nb^2(x)}$$

$\text{Bias}(\hat{m}(x)) = \frac{b^2}{2} m''(x)\mu_2(k)$
We want to show the mean and Variance of our estimator (6) using Slutsky’s theorem, but we considered the second term of (6):

\[
\hat{h}(x)\mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau)) \overset{d}{\to} h(x)\mathbb{E}(\epsilon_t | \epsilon_t > q(\tau))
\]

Let

\[
B = \hat{h}^{1/2}(x)\mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau)) \quad \text{and} \quad \mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau)) > 0
\]

Now

\[
\mathbb{E}[B] = \mathbb{E}[\hat{h}(x)\mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau))]
= \mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau))\mathbb{E}[\hat{h}(x)]
= \mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau))\left[h(x) + \frac{1}{2} \frac{\mu_2(k)}{2} \frac{h''(x)^2}{h(x)}\right]
= \mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau))h(x) + \frac{\mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau))b^2}{2} \frac{\mu_2(k)}{2} \frac{h''(x)^2}{h(x)^2}
\approx \mathbb{E}(\epsilon_t | \epsilon_t > q(\tau))h(x) + \frac{\mathbb{E}(\epsilon_t | \epsilon_t > q(\tau))b^2}{2} \frac{\mu_2(k)}{2} \frac{h''(x)^2}{h(x)^2}
\]

\[
\text{Var}[B] = \text{Var}[\hat{h}(x)\mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau))]
= \mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau))^2\text{Var}[\hat{h}(x)]
= \mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau))^2\left[\frac{1}{nbf(x)} R(k)h^2(x)\lambda^2(x)\right]
= \frac{\mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau))^2 R(k)h^2(x)\lambda^2(x)}{nbf(x)}
\approx \frac{\mathbb{E}(\epsilon_t | \epsilon_t > q(\tau))^2 R(k)h^2(x)\lambda^2(x)}{nbf(x)}
\]

Hence by Slutsky’s theorem,

\[
\hat{h}(x)\mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau)) \overset{d}{\to} h(x)\mathbb{E}(\epsilon_t | \epsilon_t > q(\tau))
\]

with mean and variance as given above.

Hence,

\[
\hat{C}\hat{E}\hat{S}(X) = \hat{m}(x) + \hat{h}(x)\mathbb{E}(\epsilon_t | \epsilon_t > \hat{q}(\tau)) = \hat{m}(x) + B = A + B
\]

Now,
\[ \mathbb{E}[\hat{CES}(X)_\tau] = \mathbb{E}[A + B] \]
\[ = \mathbb{E}[A] + \mathbb{E}[B] \]
\[ = \left[ m(x) + \frac{b^2}{2} m''(x) \mu_2(k) \right] + \mathbb{E}(\epsilon_t|\epsilon_t > \hat{q}(\tau)) \left[ h(x) + \frac{b^2}{2} \mu_2(k) h''(x)^2 \right] \]
\[ = m(x) + \frac{b^2}{2} m''(x) \mu_2(k) + \mathbb{E}(\epsilon_t|\epsilon_t > \hat{q}(\tau)) h(x) + \frac{\mathbb{E}(\epsilon_t|\epsilon_t > \hat{q}(\tau)) b^2}{2} \mu_2(k) h''(x)^2 \]
\[ \approx m(x) + \mathbb{E}(\epsilon_t|\epsilon_t > q(\tau)) h(x) + \frac{b^2}{2} \mu_2(k) \left[ m''(x) + \mathbb{E}(\epsilon_t|\epsilon_t > q(\tau)) h''(x)^2 \right] \]
\[ = \text{Bias} \]

So that,
\[ \text{Bias} \left( \hat{CES}(X)_\tau \right) = \frac{b^2}{2} \mu_2(k) \left[ m''(x) + \mathbb{E}(\epsilon_t|\epsilon_t > \hat{q}(\tau)) h''(x)^2 \right] \]
\[ \approx \frac{b^2}{2} \mu_2(k) \left[ m''(x) + \mathbb{E}(\epsilon_t|\epsilon_t > q(\tau)) h''(x)^2 \right] \]

and
\[ \text{Var} \left( \hat{CES}(X)_\tau \right) = \text{Var}(A + B) \]
\[ = \text{Var}(A) + \text{Var}(B), \quad \text{Cov}(A, B) = 0 \]
\[ = \frac{\sigma^2(x) R(k)}{nbf(x)} + \frac{\mathbb{E}(\epsilon_t|\epsilon_t > \hat{q}(\tau))^2 R(k) h^2(x) \lambda^2(x)}{nbf(x)} \]
\[ = \frac{R(k)}{nbf(x)} \left[ \sigma^2(x) + \mathbb{E}(\epsilon_t|\epsilon_t > \hat{q}(\tau))^2 h^2(x) \lambda^2(x) \right] \]
\[ \approx \frac{R(k)}{nbf(x)} \left[ \sigma^2(x) + \mathbb{E}(\epsilon_t|\epsilon_t > q(\tau))^2 h^2(x) \lambda^2(x) \right] \]

\[ \implies \hat{CES}(X)_\tau \xrightarrow{d} CES(X)_\tau, \text{ with mean and variance given above (using Slutsky’s theorem).} \]

Where

\[ \hat{CES}(X)_\tau := \mathbb{E}(Y_t|Y_t > \hat{Q}_Y|X(\tau|x), X_t = x) = \hat{m}(x) + \hat{h}^{1/2}(x) \mathbb{E}(\epsilon_t|\epsilon_t > \hat{q}(\tau)) \]

(7)

Hence, our estimator for \( CES(X)_\tau \), see [17] for its prediction intervals.
4 Smoothing Parameter (Bandwidth) Selection

In Nonparametric methods, the choice of an optimal smoothing parameter can not be over emphasized. We choose the smoothing parameter that minimizes the Asymptotic Mean Square Error (AMSE) below:

\[
AMSE(\hat{CES}(X)_{\tau}) = \mathbb{E}\left[\left(\hat{CES}(X)_{\tau} - CES(X)_{\tau}\right)^2\right]
\]

\[
= \mathbb{E}\left[\left(\hat{CES}(X) - \mathbb{E}[\hat{CES}(X)] + Bias(\hat{CES}(X))\right)^2\right]
\]

\[
= \mathbb{E}\left[\left(\hat{CES}(X) - \mathbb{E}[\hat{CES}(X)]\right)^2\right] + Bias(\hat{CES}(X)) \times \mathbb{E}\left[\hat{CES}(X) - \mathbb{E}[\hat{CES}(X)]\right] + Bias^2(\hat{CES}(X))
\]

\[
= Var(\hat{CES}(X)) + Bias^2(\hat{CES}(X))
\]

\[
= \frac{R(k)}{nb_f(x)} \left[\sigma^2(x) + \mathbb{E}(\epsilon_t|\epsilon_t > \hat{q}(\tau))h^2(x)\lambda^2(x)\right]
\]

\[
+ \frac{b^2}{2} \mu_2(k) \left[m''(x) + \mathbb{E}(\epsilon_t|\epsilon_t > \hat{q}(\tau))h''(x)^2\right]^2
\]

\[
\approx \frac{R(k)}{nb_f(x)} \left[\sigma^2(x) + \mathbb{E}(\epsilon_t|\epsilon_t > q(\tau))h^2(x)\lambda^2(x)\right]
\]

\[
+ \frac{b^4}{4} \mu_2(k) \left[m''(x) + \mathbb{E}(\epsilon_t|\epsilon_t > q(\tau))h''(x)^2\right]^2
\]

\[
(8)
\]

Therefore, \( b_{\text{opt}} = \arg\min_{b>0} AMSE(\hat{CES}(X)_{\tau}) \) and hence, \( \frac{d}{db} AMSE(\hat{CES}(X)_{\tau}) = 0 \) which gives

\[
b_{\text{opt}} = \left\{ \frac{R(k) \left[\sigma^2(x) + \mathbb{E}(\epsilon_t|\epsilon_t > q(\tau))h^2(x)\lambda^2(x)\right]}{\mu_2^2(k) f(x) \left[m''(x) + \mathbb{E}(\epsilon_t|\epsilon_t > q(\tau))h''(x)^2\right]^2} \right\}^{\frac{1}{2}} \times n^{-\frac{1}{2}}
\]

(9)
Table 1: Summary of the bootstrap results

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expection.CES</td>
<td>5.109</td>
<td>5.367</td>
<td>5.43</td>
<td>5.433</td>
<td>5.495</td>
<td>5.762</td>
</tr>
<tr>
<td>Variance.CES</td>
<td>0.732</td>
<td>1.480</td>
<td>1.7466</td>
<td>2.0631</td>
<td>2.2274</td>
<td>10.5496</td>
</tr>
<tr>
<td>Bias.CES</td>
<td>0.677</td>
<td>0.8821</td>
<td>0.9293</td>
<td>0.9371</td>
<td>0.9832</td>
<td>1.322</td>
</tr>
<tr>
<td>AMSE.CES</td>
<td>0.855</td>
<td>1.2168</td>
<td>1.3216</td>
<td>1.4031</td>
<td>1.4924</td>
<td>3.248</td>
</tr>
</tbody>
</table>

5 Simulation Study

To examine the performance of our estimators, we conducted a simulation study considering the following data generating location-scale model

\[ Y_t = m(Y_{t-1}) + h(t)^{1/2} \epsilon_t, \quad t = 1, 2, ..., n \]  \hspace{1cm} (10)

where

\[ m(Y_{t-1}) = \sin(0.5Y_{t-1}), \quad \epsilon_t \sim t(\nu = 3), \quad h(t) = h_i(Y_{t-1}) + \theta h(Y_{t-1}), \]

\[ i = 1, 2 \]

and

\[ h_1(Y_{t-1}) = 1 + 0.01Y_{t-1}^2 + 0.5\sin(Y_{t-1}), \quad h_2(Y_{t-1}) = 1 - 0.9\exp(-2Y_{t-1}^2) \]

\[ Y_t \] and \( h(t) \) are set to zero (0) initially, then \( Y_t \) is generated recursively from (10) above. The data generating process was also used by [11].

6 Application to financial market data

Using historical daily series \( \{Y_t\} \) on the log returns of the closing prices for the period between January 02, 1997 to December 29, 2017 trading days of TOTAL company quoted on the Nigerian Stock Exchange, we illustrate the applicability of the estimator.

In Figure 2 and Figure 3, the estimation of the location function and the estimation of the scale function using local linear regression are presented respectively. In Figure 1, the time series plot of the simulated daily returns is presented, Figure 4 shows the 95% conditional expected shortfall (CES) using the proposed estimator. We performed bootstrapping on the proposed estimator to verify its properties, the summary of the results is presented in Table 1. As can be seen in Table 1, the bias and asymptotic mean square error (AMSE) are relatively small, indicating a good performance of the estimator. Figure 5 and Figure 6 shows respectively the time series plot of the TOTAL’s historical market data, quoted on the Nigerian Stock Exchange and the 95% conditional expected shortfall (CES) for the TOTAL data.
Nonparametric estimation of conditional expected ...
7 Conclusion

In this paper, we have proposed a nonparametric method of estimation for conditional expected shortfall assuming that the returns on an asset or portfolio admits a location-scale model. We have verified the properties of the estimator through simulation using bootstrap method, the variance, bias, and asymptotic mean square error (AMSE) are relatively small, meaning that the estimator performs well. Financial market data from the Nigerian Stock Exchange was used for illustration.

Acknowledgements. The first author wishes to thank African Union, Pan African University, Institute for Basic Sciences Technology and Innovation Nairobi, Kenya and Japan International Cooperation (JICA) for supporting this research.

Disclosure statement. The authors declare that there is no conflict of interest regarding the publication of this paper.

References


[2] P. Artzner, F. Delbaen, J. -M. Eber and D. Heath, Thinking coherently: Generalised scenarios rather than var should be used when calculating


Received: October 7, 2019; Published: October 23, 2019