Nonparametric estimation of tail risk measures from heavy-tailed distributions

Jonthan El Methni, Laurent Gardes & Stéphane Girard

1 Tail risk measures

Let $Y \in \mathbb{R}$ be a real random loss variable. The Value-at-Risk of level $\alpha \in (0, 1)$ denoted by $\text{VaR}(\alpha)$ is the $\alpha$-quantile of the survival function $\bar{F}(x) = P(Y > x)$

$$\text{VaR}(\alpha) := \bar{F}^{-1}(\alpha) = \inf\{t, \bar{F}(t) \leq \alpha\}$$

Consider $Y_1$ and $Y_2$ two random loss variables with associated survival functions $\bar{F}_1$ and $\bar{F}_2$. 
Random variables with light tail probabilities and with heavy tail probabilities may have the same VaR(\(\alpha\)). This is one of the main criticisms against the VaR, Embrechts et al. [1997]. The Conditional Tail Expectation of level \(\alpha \in (0, 1)\) denoted by CTE(\(\alpha\)) is defined by
\[
CTE(\alpha) := \mathbb{E}(Y|Y > \text{VaR}(\alpha)).
\]

The CTE takes into account the whole information contained in the upper part of the tail distribution.
Let $Y \in \mathbb{R}$ be a real random loss variable. The Value-at-Risk of level $\alpha \in (0, 1)$ is the $\alpha$-quantile defined by

$$\text{VaR}(\alpha) := F^{-1}(\alpha),$$

where $F^{-1}$ is the (generalized) inverse of the survival function of $Y$.

The Conditional Tail Expectation of level $\alpha \in (0, 1)$ is defined by

$$\text{CTE}(\alpha) := \mathbb{E}(Y | Y > \text{VaR}(\alpha)).$$

The Conditional-Value-at-Risk of level $\alpha \in (0, 1)$ introduced by Rockafellar et Uryasev [2000] is defined by

$$\text{CVaR}_\lambda(\alpha) := \lambda \text{VaR}(\alpha) + (1 - \lambda) \text{CTE}(\alpha),$$

with $0 \leq \lambda \leq 1$.

The Conditional Tail Variance of level $\alpha \in (0, 1)$ introduced by Valdez [2005] is defined by

$$\text{CTV}(\alpha) := \mathbb{E}((Y - \text{CTE}(\alpha))^2 | Y > \text{VaR}(\alpha)).$$

The first goal of this work is to unify the definitions of the previous risk measures. To this end, the Conditional Tail Moment of level $\alpha \in (0, 1)$ is introduced:

$$\text{CTM}_a(\alpha) := \mathbb{E}(Y^a | Y > \text{VaR}(\alpha)),$$

where $a \geq 0$ is such that the moment of order $a$ of $Y$ exists. All the previous risk measures of level $\alpha$ can be rewritten as

$$\begin{align*}
\text{CTE}(\alpha) &= \text{CTM}_1(\alpha), \\
\text{CVaR}(\alpha) &= \lambda \text{VaR}(\alpha) + (1 - \lambda) \text{CTM}_1(\alpha), \\
\text{CTV}(\alpha) &= \text{CTM}_2(\alpha) - \text{CTM}_1^2(\alpha).
\end{align*}$$

All the risk measures depend on the VaR and the $\text{CTM}_a$. Our second aim is to estimate these risk measures in case of extreme losses and in the case where a covariate $X \in \mathbb{R}^p$ is recorded simultaneously with $Y$.

The fixed level $\alpha \in (0, 1)$ is replaced by a sequence $\alpha_n \to 0$.

Denoting by $F(. | x)$ the conditional survival distribution function of $Y$ given $X = x$, the Regression Value-at-Risk is defined by:

$$\text{RVaR}(\alpha_n|x) := F^{-1}(\alpha_n | x) = \inf\{t, F(t|x) \leq \alpha_n\},$$
and the Regression Conditional Tail Moment of order $a$ is defined by:

$$RCTM_a(\alpha_n|x) := \mathbb{E}(Y^a|Y > \text{RVaR}(\alpha_n|x), X = x),$$

where $a > 0$ is such that the moment of order $a$ of $Y$ exists. This yields the following risk measures:

$$\text{RCTE}(\alpha_n|x) = RCTM_1(\alpha_n|x),$$
$$\text{RCVaR}_\lambda(\alpha_n|x) = \lambda \text{RVaR}(\alpha_n|x) + (1 - \lambda)RCTM_1(\alpha_n|x),$$
$$\text{RCTV}_n(\alpha_n|x) = RCTM_2(\alpha_n|x) - RCTM_1^2(\alpha_n|x).$$

All the risk measures depend on the RVaR and the $RCTM_a$. The conditional moment of order $a \geq 0$ of $Y$ given $X = x$ is defined by

$$\varphi_a(y|x) = \mathbb{E}(Y^a I\{Y > y\}|X = x),$$

where $I\{}$ is the indicator function. Since $\varphi_0(y|x) = F(y|x)$, it follows

$$\text{RVaR}(\alpha_n|x) = \varphi_0^{-1}(\alpha_n|x),$$
$$\text{RCTM}_a(\alpha_n|x) = \frac{1}{\alpha_n} \varphi_0(\varphi_0^{-1}(\alpha_n|x)|x).$$

Goal: estimate $\varphi_a(.|x)$ and $\varphi_a^{-1}(.|x)$.

## 2 Estimators and asymptotic results

**Estimator of $\varphi_a(.,|x)$**: We propose to use a classical kernel estimator given by

$$\hat{\varphi}_{a,n}(y|x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) Y_i^a I\{Y_i > y\} / \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right).$$

- $h_n$ is a sequence called the window-width such that $h_n \to 0$ as $n \to \infty$,
- $K$ is a bounded density on $\mathbb{R}^p$ with support included in the unit ball of $\mathbb{R}^p$.

**Estimator of $\varphi_a^{-1}(.,|x)$**: Since $\hat{\varphi}_{a,n}(.,|x)$ is a non-increasing function, an estimator of $\varphi_a^{-1}(\alpha|x)$ can be defined for $\alpha \in (0, 1)$ by

$$\hat{\varphi}_{a,n}^{-1}(\alpha|x) = \inf\{t, \hat{\varphi}_{a,n}(t|x) < \alpha\}.$$
Assumptions

(F.1) The conditional survival distribution function of $Y$ given $X = x$ is assumed to be heavy-tailed i.e. for all $\lambda > 0$,

$$\lim_{y \to \infty} \frac{F(\lambda y|x)}{F(y|x)} = \lambda^{-1/\gamma(x)}.$$  

In this context, $\gamma(\cdot)$ is a positive function of the covariate $x$ and is referred to as the conditional tail index since it tunes the tail heaviness of the conditional distribution of $Y$ given $X = x$. Condition (F.1) also implies that for $a \in [0, 1/\gamma(x))$, RCTM$_a(\cdot|x)$ exists, and for all $y > 0$,

$$\text{RCTM}_a(1/y|x) = y^{a\gamma(x)} \ell_a(y|x),$$

where for $x$ fixed, $\ell_a(\cdot|x)$ is a slowly-varying function i.e. for all $\lambda > 0$,

$$\lim_{y \to \infty} \frac{\ell_a(\lambda y|x)}{\ell_a(y|x)} = 1.$$

(F.2) $\ell_a(\cdot|x)$ is normalized for all $a \in [0, 1/\gamma(x))$.

In such a case, the Karamata representation of the slowly-varying function can be written as

$$\ell_a(y|x) = c_a(x) \exp \left( \int_1^y \frac{\varepsilon_a(u|x)}{u} du \right),$$

where $c_a(\cdot)$ is a positive function and $\varepsilon_a(y|x) \to 0$ as $y \to \infty$.

(F.3) $|\varepsilon_a(\cdot|x)|$ is continuous and ultimately non-increasing for all $a \in [0, 1/\gamma(x))$.

A Lipschitz condition on the probability density function $g$ of $X$ is also required:

(L) There exists a constant $c_g > 0$ such that $|g(x) - g(x')| \leq c_g d(x, x')$.

where $d(x, x')$ is the Euclidean distance between $x$ and $x'$. Finally, for $y > 0$ and $\xi > 0$, the largest oscillation of the conditional moment of order $a \in [0, 1/\gamma(x))$ is defined by

$$\omega(x, y, a, \xi, h_n) = \sup \left\{ \left| \frac{\varphi_a(z|x)}{\varphi_a(z|x')} - 1 \right| : z \in [(1 - \xi) y, (1 + \xi) y], x' \in B(x, h_n) \right\},$$

where $B(x, h_n)$ denotes the ball centred at $x$ with radius $h_n$.  

5
Theorem 1 Suppose (F.1), (F.2) and (L) hold. Let

- $0 \leq a_1 < a_2 < \cdots < a_J$,
- $x \in \mathbb{R}^p$ such that $g(x) > 0$ and $0 < \gamma(x) < 1/(2a_J)$,
- $\alpha_n \to 0$ and $nh_n^p\alpha_n \to \infty$ as $n \to \infty$,
- $\xi > 0$ such that $\sqrt{nh_n^p\alpha_n}(h_n \vee \max_{a_1} \omega(x, \text{RVaR}(\alpha_n|x, a, \xi, h_n))) \to 0$.

Then,

$$
\sqrt{nh_n^p\alpha_n} \left\{ \left( \frac{\text{RCTM}_{a_{j,n}}(\alpha_n|x)}{\text{RCTM}_{a_j}(\alpha_n|x)} - 1 \right)_{j \in \{1, \ldots, J\}}, \left( \frac{\text{RVaR}_{n}(\alpha_n|x)}{\text{RVaR}(\alpha_n|x)} - 1 \right) \right\}
$$

is asymptotically Gaussian, centered, with covariance matrix $\| K \|^2 \gamma^2(x) \Sigma(x)/g(x)$ where

$$
\Sigma(x) = \begin{pmatrix}
  a_1 & 1 \\
  \vdots & \vdots \\
  a_J & a_1 \cdots a_J \\
  1 & a_1 \cdots a_J
\end{pmatrix}.
$$

**Conditions on the sequences $\alpha_n$ and $h_n$.**

$nh_n^p\alpha_n \to \infty$: Necessary and sufficient condition for the almost sure presence of at least one point in the region $B(x, h_n) \times [\text{RVaR}(\alpha_n|x), +\infty)$ of $\mathbb{R}^p \times \mathbb{R}$.

$$
\sqrt{nh_n^p\alpha_n}(h_n \vee \max_{a_1} \omega(x, \text{RVaR}(\alpha_n|x, a, \xi, h_n))) \to 0:
$$

The bias induced by the smoothing is negligible compared to the standard-deviation.
Corollary 1 Suppose the assumptions of Theorem 1 hold. Then, if $0 < \gamma(x) < 1/2$,

\[
\sqrt{nh_n^p} \left( \frac{\widehat{\text{RCTE}}_{n}(\alpha_n|x) - 1}{\text{RCTE}(\alpha_n|x)} \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{2(1 - \gamma(x))\gamma^2(x) \|K\|_2^2}{1 - 2\gamma(x)g(x)} \right)
\]

\[
\sqrt{nh_n^p} \left( \frac{\widehat{\text{RCVaR}}_{\lambda,n}(\alpha_n|x) - 1}{\text{RCVaR}_{\lambda}(\alpha_n|x)} \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\gamma^2(x)(\lambda^2 + 2 - 2\lambda - 2\gamma(x)) \|K\|_2^2}{1 - 2\gamma(x)g(x)} \right)
\]

The $\text{RCTV}(\alpha_n|x)$ estimator involves the computation of a second order moment, it requires the stronger condition $0 < \gamma(x) < 1/4$,

\[
\sqrt{nh_n^p} \left( \frac{\widehat{\text{RCTV}}_{n}(\alpha_n|x) - 1}{\text{RCTV}(\alpha_n|x)} \right) \xrightarrow{d} \mathcal{N} \left( 0, V_{\gamma(x)} \frac{\|K\|_2^2}{g(x)} \right),
\]

where

\[
V_{\gamma(x)} = \frac{8(1 - \gamma(x))(1 - 2\gamma(x))(1 + 2\gamma(x) + 3\gamma^2(x))}{(1 - 3\gamma(x))(1 - 4\gamma(x))}.
\]

3 Illustration on a simulated dataset

A sample $\{(X_i, Y_i), i = 1, \ldots, n\}$ of size $n = 1000$ is generated. The covariate $X$ is uniform on $[0, 1]$. The conditional distribution of $Y|X = x$ is chosen in the Hall class:

\[
F(y|x) = y^{-1/\gamma(x)} a(1 + b y^{\rho/\gamma(x)}) \ell(y|x)
\]

with $a = 1/2$, $b = 1$, $\rho = -1$ and conditional tail index function

\[
x \in [0, 1] \to \gamma(x) = \frac{1}{2} \left( \frac{1}{10} + \sin(\pi x) \right) \left( \frac{11}{10} - \frac{1}{2} \exp \left( 64 \left( x - \frac{1}{2} \right)^2 \right) \right).
\]

We have chosen a bi-quadratic kernel

\[
K(u) \propto (1 - u^2)^2 \mathbb{1}_{\{|u| \leq 1\}}
\]

with smoothing parameter $h_n = 0.1$. Our goal if to estimate $\text{RCTE}(\alpha_n|x)$ and $\text{RCVaR}(\alpha_n|x)$ with $\alpha_n = 0.05$. 

7
Theoretical RCTE($\alpha_n|x$) and RVaR($\alpha_n|x$) with a logarithmic scale.

Theoretical and estimated RVaR($\alpha_n|x$) with a logarithmic scale.
Theoretical and estimated $\text{RCTE}(\alpha_n|x)$ with a logarithmic scale.
Theoretical and estimated RCTE(α_n|x) with a logarithmic scale.

4 Extrapolation: A Weissman type estimator

- In Theorem 1, the condition \( nh_n^p \alpha_n \to \infty \) provides a lower bound on the level of the risk measure to estimate.

- This restriction is a consequence of the use of a kernel estimator which cannot extrapolate beyond the maximum observation in the ball \( B(x, h_n) \).

- In consequence, \( \alpha_n \) must be an order of an extreme quantile within the sample.

Let us consider \((\alpha_n)_{n \geq 1}\) and \((\beta_n)_{n \geq 1}\) two positive sequences such that \( \alpha_n \to 0, \beta_n \to 0 \) and \( 0 < \beta_n < \alpha_n \). A kernel adaptation of Weissman’s estimator [1978] is given by

\[
\text{RCTM}_{a,n}^W(\beta_n|x) = \text{RCTM}_{a,n}(\alpha_n|x) \left( \frac{\alpha_n}{\beta_n} \right)^{a \hat{\gamma}_n(x)} \text{extrapolation}
\]

**Theorem 2** Suppose the assumptions of Theorem 1 hold together with (F.3). Let \( \hat{\gamma}_n(x) \) be an estimator of the conditional tail index such that

\[
\sqrt{nh_n^p \alpha_n} (\hat{\gamma}_n(x) - \gamma(x)) \overset{d}{\to} \mathcal{N} \left( 0, v^2(x) \right),
\]
with \( v(x) > 0 \). If, moreover \((\beta_n)_{n \geq 1}\) is a positive sequence such that \( \beta_n \to 0 \) and \( \beta_n/\alpha_n \to 0 \) as \( n \to \infty \), then

\[
\frac{\sqrt{n h_n^p \alpha_n}}{\log(\alpha_n/\beta_n)} \left( \frac{\widehat{RCTM}_{a,n}(\beta_n|x)}{RCTM_a(\beta_n|x)} - 1 \right) \xrightarrow{d} N(0, (\mathbb{E}v(x))^2).
\]

The condition \( \beta_n/\alpha_n \to 0 \) allows us to extrapolate and choose a level \( \beta_n \) arbitrarily small.

### Estimation of the conditional tail index

- **Without covariate:** Hill [1975]. Let \((k_n)_{n \geq 1}\) be a sequence of integers such that \( k_n \in \{1 \ldots n\} \). The Hill estimator is given by

\[
\hat{\gamma}_{n,\alpha_n} = \frac{1}{k_n - 1} \sum_{i=1}^{k_n - 1} \log \frac{Z_{n-i+1,n}}{Z_{n-k_n+1,n}},
\]

where \( Z_{1,n} \leq \cdots \leq Z_{n,n} \) are the order statistics associated with i.i.d. random variables \( Z_1, \ldots, Z_n \) and \( \alpha_n = k_n/n \).

- **With a covariate:** A kernel version of the Hill estimator is given by

\[
\hat{\gamma}_{n,\alpha_n}(x) = \frac{\sum_{j=1}^{J} (\log \widehat{RVaR}_n(\tau_j \alpha_n|x) - \log \widehat{RVaR}_n(\tau_1 \alpha_n|x))}{\sum_{j=1}^{J} \log(\tau_1/\tau_j)},
\]

where \( J \geq 1 \) and \((\tau_j)_{j \geq 1}\) is a decreasing sequence of weights.

The asymptotic normality of \( \hat{\gamma}_{n,\alpha_n}(x) \) has been established by Daouia et al. [2011]. As a consequence, replacing \( \widehat{RVaR} \) by \( \widehat{RVaR}_n^W \) and \( RCTM_{a,n} \) by \( \widehat{RCTM}_{a,n}^W \) provides (asymptotically Gaussian) estimators for all the risk measures considered here, and for arbitrarily small levels. In particular, since \( \widehat{RCTE}(\alpha_n|x) = RCTM_1(\alpha_n|x) \), we obtain

\[
\widehat{RCTE}_n^W(\beta_n|x) = \widehat{RCTE}_n(\alpha_n|x) \left( \frac{\alpha_n}{\beta_n} \right)^{\hat{\gamma}_n(x)}.
\]
References


