MINERVA INVESTMENT MANAGEMENT SOCIETY
Quantitative Research Division

Coherent Risk Measures

Umberto Barbieri	Head of Quantitative Research
Antonio Petrai	Quantitative Research Analyst
Enrico Sammarco	Quantitative Research Analyst
Erik Vik	Quantitative Research Analyst

Academic Year 2023/2024

Abstract

In this paper we introduce the notion of Coherent Risk Measures and the desirable mathematical properties that they satisfy. We focus on the Conditional VaR (CVaR) and provide a framework for its estimation that is consistent with Extreme Value Theory. We conclude by presenting the Entropic VaR (EVaR) as an alternative risk measure that, while retaining coherency, is also computationally tractable.

1 Introduction

Risk measures are related to the quantification of the risk associated with a financial position. To characterize such position, we rely on the corresponding payoff profile, represented by a real-valued random variable X belonging to a given class $\mathcal{X}$ of well-behaved random variables defined over an appropriate probability space $(\Omega, \mathcal{F}, P)$.

Under a probabilistic model, the risk inherent in X can be assessed by considering statistical quantities like moments and quantiles of the probability distribution of scenarios. However, straightforward metrics like variance fail to account for a fundamental asymmetry in the financial interpretation of X, where the **downside risk** holds more significance. This discrepancy is addressed by quantities that focus on the left tail of the payoff distribution. Nonetheless, the widely known Value at Risk (VaR) falls short in meeting some essential consistency criteria, motivating a more systematic exploration of risk measures and their "desirable" properties.

If we adopt, for instance, the perspective of a regulatory agency a risk measure is treated as a capital requirement, which determines the minimum amount of capital that, when added to the financial position and invested risk-free, renders the position "acceptable". This monetary interpretation of the financial position is encapsulated by an additional property known as translation (or cash) invariance. When combined with convexity and monotonicity, this leads to the definition of coherent risk measures [3].

Coherent Risk Measure

Definition 1 *A mapping $\rho : \mathcal{X} \to \mathbb{R}$ is called a coherent measure of risk if it satisfies the following conditions for all $X_1, X_2 \in \mathcal{X}$ and for $\lambda \in [0, 1]$:*

- ***Monotonicity:*** *If $X_1 \leq X_2$, then $\rho(X_1) \leq \rho(X_2)$*
- ***Translation Invariance:*** *If $m \in \mathbb{R}$, then $\rho(X + m) = \rho(X) - m$.*
- ***Convexity:*** $\rho(\lambda X_1 + (1 - \lambda)X_2) \leq \lambda\rho(X_1) + (1 - \lambda)\rho(X_2)$
- ***Positive Homogeneity:*** $\rho(\lambda X) = \lambda\rho(X)$
- ***Subadditivity:*** $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$

Notably, a pivotal result in the field is that coherent risk measures admit a **dual representation** and can be characterized as:

$$\rho(X) = \sup_{Q \in \mathbb{Q}} E^Q[X] \tag{1}$$

where $\mathbb{Q}$ is some particular class of probability measures on the sample space, whose representation is specific to each risk measure [6]. It is clear that, once having picked a suitable member $X \in \mathcal{X}$ and if we prefer smaller values of X (e.g., if X represents a distribution of losses), then we are interested in the following optimization problem to make the selection:

$$\min_{X \in \mathcal{X}} \rho(X). \tag{2}$$

We will tackle again the problem in the section dedicated to EVaR.

2 Conditional VaR

2.1 Overcoming VaR limitations

The **Value at Risk** (VaR) is a well known and widely implemented risk measure that indicates the loss amount that is exceeded only with a small probability over the horizon of interest (usually a day or a year). Formally, for a confidence level $1-\alpha$ the $VaR^{\alpha}(X)$ is the α-quantile of the payoff distribution for a financial position X, i.e.

$$VaR^{\alpha}(X) = \inf_{x\in\mathbb{R}}\{x : P(X \leq x) \geq 1-\alpha\} \tag{3}$$

As we have hinted to in the previous section, VaR is not a coherent risk measure. Indeed, it lacks subadditivity, i.e. given two random variables X_1 and X_2 it is not generally guaranteed that

$$VaR(X_1 + X_2) \leq VaR(X_1) + VaR(X_2) \tag{4}$$

Other shortcomings of the VaR are that, for **non-normal distributions**, working numerically with the VaR is unstable and optimization becomes **intractable in higher dimensions**. Moreover, from a risk management perspective, VaR does not control scenarios and losses exceeding VaR itself.

The research for a more amenable risk measure lead to the definition of **Conditional Value at Risk** (CVaR). The CVaR is also known in the financial milieu with the name Expected Shortfall (ES): even if these two metrics do not exactly coincide, they both refer to the expected size of a loss X exceeding the VaR, i.e.

$$ES^{\alpha}(X) = E[X|X > VaR^{\alpha}(X)] \tag{5}$$

$$CVaR^{\alpha}(X) = \frac{1}{\alpha}\int_0^{\alpha} VaR^{a}(X)\, da \tag{6}$$

For simplicity and with a slight abuse of notation we shall consider the two as equivalent. As opposed to VaR, the **CVaR is a coherent risk measure** for general loss distribution (see [10]). Moreover, [2] show that it admits the dual representation of Equation 1.

2.2 EVT modeling framework

In this section we outline a framework for the estimation of the CVaR of a financial position that is consistent with Extreme Value Theory (EVT). Broadly spekaing, EVT is a branch of statistics that deals with events that significantly deviate from the median of a distribution. It provides a mathematically rigorous approach to estimate and extrapolate the probability of extreme events that can be hardly observed in a dataset.

From a financial point of view, we restrict the analysis to *market risk*. The type of losses we have in mind are those due to adverse market movements while holding a position on a single or a portfolio of traded securities. From a modeling perspective it is also convenient to leave aside the monetary dimension of X and to focus our attention on (log) returns.

In fact, the problem of computing the CVaR of a p&l with initial position W_0 is equivalent to finding the CVaR of the return r distribution, since the mapping $r \rightarrow$ X is both continuous and strictly increasing:

$$X = W_1 - W_0 = W_0(e^r - 1). \tag{7}$$

EVT results are usually stated with the aim of modeling the maximum or the right tail of a distribution. To adapt them to our setting we need to consider the distribution of **negative log-returns** $-r \equiv r$ and then translate our conclusions back to the original distribution in a second moment.

In order to properly apply most of the statistical results in (univariate) EVT we need to handle random variables that are **approximately i.i.d.** (see [5]). The assumption of return i.i.d.ness can be particularly strong when modeling traded securities as it is not consistent with many stylized facts (e.g. volatility clustering). Moreover, for risk management purposes, it is useful to have some model to make good use of the information we have and thus makes conditional forecasts rather than unconditional ones. These considerations justify the assumption of a **data generating process** for the returns of the form

$$r_{t+1|t} = \mu_{t+1|t} + \sigma_{t+1|t} Z_{t+1} \tag{8}$$

for $t = 0, 1, ..., T$, where $\mu_{t+1|t}$ and $\sigma_{t+1|t}$ are some model specifications for the conditional mean and volatility, respectively (e.g. ARMA-GARCH, see [8] for a more in depth description). If the other components have been correctly specified, we can assume the Z_t (commonly referred as *shocks*) to be i.i.d. from a generic standardised distribution F.

The useful reflection of 8 is that we can apply the EVT toolkit to the filtered residuals to get an estimate of VaR and CVaR for the the returns:

$$\begin{cases} VaR^{\alpha}_{t+1|t} = \mu_{t+1|t} + \sigma_{t+1|t} q^{\alpha}(Z) \\ CVaR^{\alpha}_{t+1|t} = \mu_{t+1|t} + \sigma_{t+1|t} CVaR^{\alpha}(Z) \end{cases} \tag{9}$$

for a given level α and $q_{\alpha}(\cdot) := \inf\{x : F(x) \geq 1 - \alpha\}$ being the quantile function. It is worth noticing that, due to i.i.d.ness, $q^{\alpha}(Z)$ and $CvaR^{\alpha}(Z)$ are not time-dependent.

The next step is where EVT comes into play. We adopt the **Peaks Over Threshold** (POT) approach to model the exceedances above a threshold $u \in \mathbb{R}$ high enough, that is we consider $Z - u \mid Z > u$ with CDF

$$F_u(z) = P(Z - u \leq z \mid Z > u) = \frac{F(z+u) - F(u)}{1 - F(u)} \tag{10}$$

where F_u is also called *excess distribution function*.

In practical settings, however, the distribution F_u is **unknown**. We therefore advocate for some limiting results that allow to approximate F_u with a known distribution G. A central theorem in EVT from [4] and [9] offers a way out by specifying the functional form of G as the threshold u approaches the right endpoint of the distribution z^*.

Pickands–Balkema–de Haan

Theorem 2.1 *Let F_u be an excess distribution function for a threshold u, then $\forall \xi \in \mathbb{R}$ we have $F \in \mathcal{D}(G(\xi, \beta(u)))$ if and only if*

$$\lim_{u \to z^*} \sup_{0 \le z \le z^* - u} |F_u(z) - G(z; \xi; \beta(u))| = 0 \tag{11}$$

where $\mathcal{D}$ denotes the maximum domain of attraction and

$$G(z; \xi, \beta(u)) = \begin{cases} 1 - (1 + \xi \frac{z}{\beta(u)})^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp(-\frac{z}{\beta(u)}) & \xi = 0 \end{cases} \tag{12}$$

is a Generalized Pareto distribution with scale $\beta(u) > 0$.

The Generalized Pareto distribution (GPD) subsumes other distributions under a common parametric form, where the parameter ξ influences the *shape* of the distribution. In our setting, two cases are relevant:

- $\xi > 0$, which is indicative of **heavy-tailed** behavior and power law decay; F is said to be in the domain of attraction of the Fréchet distribution,
- $\xi = 0$, which is typical of **thin-tailed** distributions and exponential decay; moreover F is in the domain of attraction of the Gumbel distribution.

Under this framework we can approximate the CDF of Z with

$$\begin{aligned} F(z+u) &= F(u) + (1 - F(u)) \frac{F(z+u) - F(u)}{1 - F(u)} \\ &\approx F(u) + (1 - F(u)) \left(1 - \left(1 + \xi \frac{z}{\beta(u)}\right)^{-\frac{1}{\xi}}\right) \end{aligned} \tag{13}$$

as long as u is high enough and for $\xi \neq 0$.

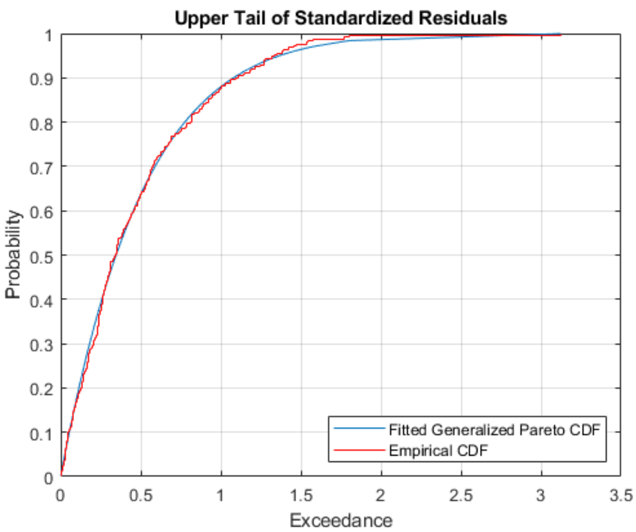

Figure 1: GPD fitted to the upper tail of standardized residuals vs their empirical CDF.

Equation 13 can be inverted to obtain closed-form expressions for both the VaR and the CVaR for Z.

Closed-form expressions for $VaR^{\alpha}(Z)$ and $CVaR^{\alpha}(Z)$

As $u \to z^*$ and for $0 < \xi < 1$,

$$VaR^{\alpha}(Z) = q^{\alpha}(Z) = u + \frac{\beta(u)}{\xi}\left(\left(\frac{\alpha}{1-F(u)}\right)^{-\xi} - 1\right) \tag{14}$$

$$CVaR^{\alpha}(Z) = \frac{VaR^{\alpha}(Z)}{1-\xi} + \frac{\beta(u) - \xi u}{1-\xi} \tag{15}$$

Similar expressions hold for $\xi = 0$. We can ultimately plug these equations in 9 to get a conditional estimate for the VaR and CVaR of r given $\mu_{t+1|t}$ and $\sigma_{t+1|t}$.

2.3 Parameters estimation

Now that a framework has been set, we are left with the nontrivial task of estimating the parameters of the GPD, namely ξ, $\beta(u)$ and the CDF evaluated at the threshold $F(u)$. Here we briefly present the main approaches that can be followed, for an in depth analysis of the estimation process the interested reader shall see [7].

Under the POT approach the k top order statistics are considered in the estimation process, where $k \to \infty$ as $n \to \infty$ and $k = o(n)$. The choice of k (known as *effective sample size*) itself entails a **bias-variance trade-off**:

- A very high k and a low threshold increase the bias we get in the estimation of the other parameters as the approximation in Theorem 2.1 becomes poorer,
- Low values of k and a higher threshold help containing the bias, but increase the variance of our estimates as we are working with less observations.

Given k, a natural way to proceed is to set u equal to some value in the observed dataset, usually the $(n-k)$th ordered statistic X_{n-k} and to estimate $F(u)$ with the empirical CDF $\hat{F}(u)$. On the other hand, common methods to estimate ξ and $\beta(u)$ include:

- **Maximum Likelihood Estimation** (MLE), which jointly estimates the two parameters by maximizing the log-likelihood of the observed exceedances under the GPD approximation,
- **Moment Estimator**, which leads to non-parametric estimators for both ξ and $\beta(u)$; in the case of ξ the moment estimator is

$$\hat{\xi_M} := M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right)^{-1} \tag{16}$$

where $M_n^{(j)}$ is defined as

$$M_n^{(j)} := \frac{1}{k}\sum_{i=0}^{k-1}(\log X_{n-i} - \log X_{n-k})^j, \tag{17}$$

- **Bayesian approach**, which is not restricted to a closed-form estimator but tries to deliver the joint posterior distribution for ξ and $\beta(u)$ while accounting for some prior information.

As soon as we have a point estimate for $F(\hat{u})$, $\hat{\xi}$ and $\beta(\hat{u})$ we can plug everything back in Equations 14 and 15 to get a point estimate for VaR and CVaR of Z.
Under second-order conditions outlined by [7], the MLE and Moment estimators can be shown to be consistent and asymptotically normal, but still biased. These properties turn out to be particularly useful in quantifying the uncertainty around our estimates as well as in performing some sensitivity analysis w.r.t. the chosen threshold u (and k).

3 Entropic VaR

By following the seminal work [1], we introduce a recently developed coherent risk measure: the Entropic Value at Risk (EVaR).
EVaR appeared in the literature as an attempt to overcome the drawbacks of its popular predecessors VaR and CVaR. In Section 2.1 we have shown that VaR lacks coherency and it is computationally intractable when employed in optimization problems. In Section 2.2 we have outlined how the CVaR can be estimated for risk management purposes; however, also CVaR poses computational challenges when used in optimization, even in the "simple" summation of independent random variables. Efficient computation of CVaR is limited to a few specific cases, requiring approximation through sampling methods for more complex scenarios. Having a risk measure that can be computed efficiently becomes crucial when the dimensionality of the problem escalates quickly.

The EVaR is obtained as the tightest possible upper bound obtained from the **Chernoff inequality** for the VaR

$$\Pr(X \geq a) \leq e^{-za} M_X(z) \tag{18}$$

$\forall a \in \mathbb{R}$ and $\forall z > 0$, where $M_X(z) = E[e^{zX}]$ is the moment-generating function of X.

Entropic VaR

Definition 2 *The Entropic VaR of X with confidence level $1-\alpha$ is*

$$EVaR^{\alpha}(X) = \inf_{z>0} \left\{ \frac{1}{z} \ln \left(\frac{M_X(z)}{\alpha} \right) \right\} \tag{19}$$

It can be shown that the $EVaR^{\alpha}$ is coherent for every $\alpha \in (0,1]$ and admits the dual representation $EVaR^{\alpha}(X) = \sup_{Q \in \mathbb{Q}} E^Q[X]$.

As an example, we can compare the closed formula expressions for VaR, CVaR, and EVaR when $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$VaR^{\alpha}(X) = \mu + z_{1-\alpha}\sigma, \tag{20}$$

$$CVaR^{\alpha}(X) = \mu + \frac{\phi(z_{1-\alpha})}{\alpha}\sigma, \tag{21}$$

$$EVaR^{\alpha}(X) = \mu + \sqrt{-2\ln\alpha}\,\sigma \tag{22}$$

where $\phi(\cdot)$ is the density function and $z_{1-\alpha}$ is such that $P(Z \geq z_{1-\alpha}) = \alpha$.

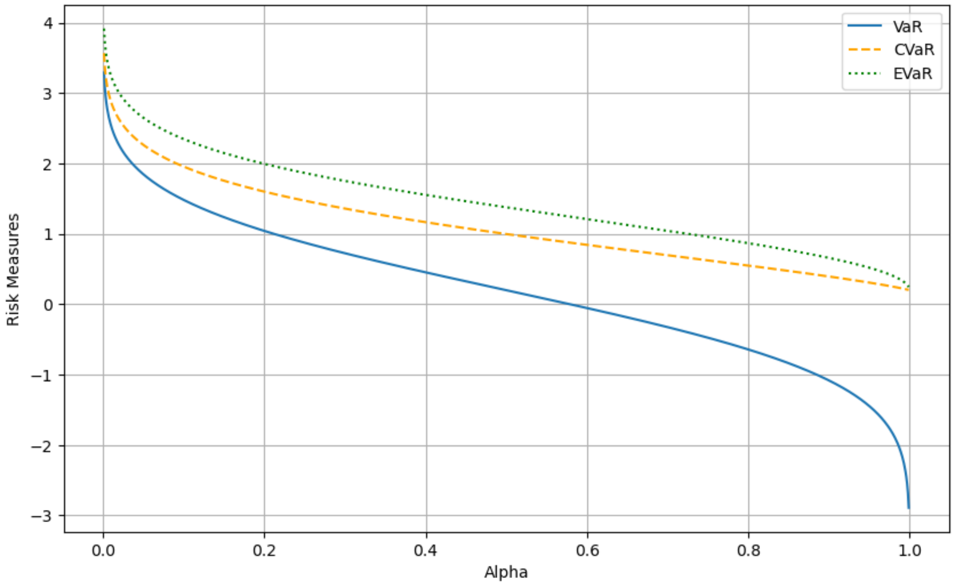

Figure 2: VaR, CVaR and EVaR for normally distributed random variables.

We are now going to focus on the following proposition, which gives a characterization of EVaR as an upper bound for both VaR and CVaR.

EVaR as an upper bound

Proposition 3.1 *The EVaR is an upper bound for both the VaR and the CVaR with the same confidence levels, i.e., for X and every $\alpha \in (0, 1]$*

$$VaR^{\alpha}(X) \leq CVaR^{\alpha}(X) \leq EVaR^{\alpha}(X) \tag{23}$$

This proposition establishes that the EVaR exhibits a higher degree of risk aversion in comparison to the CVaR when considering the same confidence level. Consequently, the EVaR induces an allocation of resources more tilted towards risk mitigation. However, this fact may not align with the preferences of companies aiming to minimize resource allocation. This particular aspect diminishes the attractiveness of the EVaR for such companies. Nonetheless, its most noteworthy attribute lies in its computational feasibility for numerous scenarios where the CVaR faces challenges. In fact, when the incorporation of a risk measure into a stochastic optimization problem is necessary, the computational tractability of the EVaR becomes a vital consideration.

3.1 EVaR in optimization

In practical applications, composite random variables, denoted as $X = H(\mathbf{w}, \mathbf{\Psi})$, are frequently employed. Here, $\mathbf{w}$ represents an n-dimensional real decision vector in $\mathbf{W} \subseteq \mathbb{R}^n$, $\mathbf{\Psi}$ is an m-dimensional real random vector with a known probability distribution, and the function $H(\mathbf{w}, \cdot) : \mathbb{R}^m \to \mathbb{R}$ is a measurable function for all $\mathbf{w} \in \mathbf{W}$. Consequently, the original problem in 2 can be expressed as:

$$\min_{\mathbf{w} \in \mathbf{W}} \rho(H(\mathbf{w}, \mathbf{\Psi})) \tag{24}$$

It can be shown that the minimization problem is a **convex** if the risk measure ρ is coherent and if the function $H(\cdot, \mathbf{s})$ is convex for all $\mathbf{s}$ in $\mathbf{S}_{\boldsymbol{\Psi}}$, where $\mathbf{S}_{\boldsymbol{\Psi}}$ represents the support of the random vector $\boldsymbol{\Psi}$, that is if the function is convex in the vector-values $\boldsymbol{\Psi}$ can assume. If ρ is then taken to be the EVaR, the problem becomes:

$$\min_{\mathbf{w}\in\mathbf{W}, z>0} \left\{ \frac{1}{z} \ln \left(\frac{M_{H(\mathbf{w},\boldsymbol{\Psi})}(z)}{\alpha} \right) \right\} \tag{25}$$

For instance, in a simple portfolio setting where the choice variable $\mathbf{w}$ represents the individual securities' weights, the random vector $\boldsymbol{\Psi}$ is the vector of random linear returns $R = e^r - 1$ for each single security, and $H(\cdot, \mathbf{s})$ is the linear return of the portfolio, the optimization problem employing EVaR as a risk measure becomes:

$$\min_{\mathbf{w}\in\mathbf{W}} \left\{ EVaR^{\alpha} \left(\sum_{i=1}^{n} w_i R_i \right) \right\} \tag{26}$$

where $H(\mathbf{w}, \boldsymbol{R}) = \sum_{i=1}^{n} w_i R_i$ is clearly affine in $\boldsymbol{R}$. In [1] it is also shown that $H(\cdot, \mathbf{s})$ is convex for all $\mathbf{s}$ in $\mathbf{S}_{\boldsymbol{R}}$.
The problem becomes computationally tractable if the objective function can be computed efficiently, e.g., if $M_{H(\mathbf{w},\boldsymbol{\Psi})}(z)$ can be evaluated in polynomial time, but most importantly it becomes entirely tractable if $M_{H(\mathbf{w},\boldsymbol{\Psi})}(z)$ is affine in $\boldsymbol{\Psi}$.

Moreover, to give a concrete instance of the reduced computational complexity of the problem we provide the proposition below, as seen in [1].

Computational complexity of EVaR (discrete case)

Proposition 3.2 *Let $\Psi_1, \ldots, \Psi_m$ be independent discrete random variables assuming k distinct values, and $H(\mathbf{w}, \boldsymbol{\Psi})$ be affine in $\boldsymbol{\Psi}$ for all $\mathbf{w} \in \mathbf{W}$. For fixed $\mathbf{w}$ and z, the computational complexity of the objective function in the minimization problem, when using EVaR, is a bilinear function of m and k (mk), while for CVaR, it is of order k^m, growing exponentially with m and polynomially with k.*

We conclude with a remark regarding the reason behind the adjective "entropic". Until now we have not specified what the class $\mathbb{Q}$ of probability measures in the dual representation looks like. [1] shows that, in the case of EVaR,

$$\mathbb{Q} = \{Q \ll P : D_{KL}(Q \parallel P) \leq -\ln \alpha\}$$

where $D_{KL}(Q \parallel P)$ is the Kullback–Leibler divergence from Q to P, or more distinctly, the **relative entropy** of Q with respect to P. This divergence intuitively measures how much the probability measure Q is different ("distant") from the probability measure P and is defined as the Q-expectation of the logarithmic difference between the two measures:

$$D_{KL}(Q \parallel P) = \int_{\Omega} \ln \left(\frac{Q(d\omega)}{P(d\omega)} \right) Q(d\omega) \tag{27}$$

where $Q(d\omega)/P(d\omega)$ is the Radon-Nikodym derivative of Q with respect to P.

References

[1] A Ahmadi-Javid. "Entropic Value-at-Risk: A New Coherent Risk Measure". In: *Journal of Optimization Theory and Applications* 155 (2012), pp. 1105–1123. DOI: https://doi.org/10.1007/s10957-011-9968-2.

[2] M. Ang, J. Sun, and Q. Yao. "On the dual representation of coherent risk measures". In: *Annals of Operations Research* 262 (2018), 29:46. DOI: 10.1007/s10479-017-2441-3.

[3] P. Artzner et al. *Coherent Measures of Risk*. 1999. DOI: https://doi.org/10.1111/1467-9965.00068.

[4] A. A. Balkema and L. de Haan. "Residual Life Time at Great Age". In: *The Annals of Probability* 2.5 (1974), pp. 792–804. DOI: 10.1214/aop/1176996548.

[5] Stuart Coles. *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag, 2001. DOI: https://doi.org/10.1007/978-1-4471-3675-0.

[6] Föllmer H. and A. Schied. *Stochastic Finance An Introduction in Discrete Time*. Walter de Gruyter, Berlin · New York, 2004. DOI: https://doi.org/10.1515/9783110212075.

[7] L. de Haan and A. Ferreira. *Extreme Value Theory*. Springer New York, NY, 2006. DOI: https://doi.org/10.1007/0-387-34471-3.

[8] J. Hamilton. *Time Series Analysis*. Princeton University Press, 1994. DOI: https://doi.org/10.2307/j.ctv14jx6sm.

[9] James Pickands III. "Statistical Inference Using Extreme Order Statistics". In: *The Annals of Statistics* 3.1 (1975), pp. 119–131. DOI: 10.1214/aos/1176343003.

[10] R. T. Rockafellar and S. Uryasev. "Conditional value-at-risk for general loss distributions". In: *Journal of Banking Finance* 26 (2002), 1443:1471. DOI: https://doi.org/10.1016/S0378-4266(02)00271-6.