Risk Measures and Robustness

2016-2017 CAE Actuarial Science Research Project

Supervised by Dr. Ying Wang

Jinkai Xu
Nur Amalina Abdul Razak
Charlies W. Robinson
Peng Jin

Department of Mathematics

University of Illinois at Urbana-Champaign

May 20, 2017
Abstract

In this paper, we will summarize the risk measures and their robustness metrics. Simulate and perform research on the robustness for VaR and TVaR.

Keywords: Risk measures, robustness, sensitivity, VaR, TVaR
1 Introduction

Measuring or quantifying risk is important to understand the potential features of risk that an institution has. It helps to analyze the efficiency of risk control measures which is significant in the process of decision making. However, this attempt sounds controversial as it tries to quantify the whole statistical distribution of financial loss with a single number (Kuo et al., 2010). Currently, there are many types of risk measures that have been developed to accommodate this situation. Most of these precedent risk measures are favored to hold the property of coherence, but this modeling assumption is too sensitive for the tails of loss distributions and outliers. This sensitivity is known as robustness; which is an essential feature for a risk measure, especially for regulatory purposes. Otherwise, Kuo (2013) stated that regulatory risk measures would be unacceptable because different regulatory capital needs are specific to each institution.

The robustness of policy rules pertains to the property of well performing across of different of the alternative model including integrates the misspecification errors because it has a close correlation to ambiguity aversion and model uncertainty (Kuo et al., 2010). This is because risk measures have two objectives: the internal objective for individual institution risk management and the external objective for all relevant external institution regulations. Kuo et al. (2010) signified that the difference depends on how much information is available to tailor the risk measure. The paper also added that our current regulation has allowed the usage of internal modeling and private data with external regulations. This has lead to two points at issue which are unreliable data and attaining several models for the same portfolio. Thus, other than having an external risk measure that demonstrates societal norms, it should be robust along with balancing sub-additivity to allow comparisons between different distortion functions or probability measures. For instance, Kuo (2013) implied that using the median of a distribution would produce a better robust measurement because it considers the size of a position and the likelihood of losses when evaluating a specific risk. In addition, there are more concerns revolving around this topic such as robustness and conservative risk measures are preferable by regulators of rigidness and diversification would rely upon the tail of the distribution (Kuo et al., 2010).
2 Risk Measures

In this section, we will introduce various categories for risk measures. The first category is defining the general types of risk measures and their mathematical properties that a risk measure must satisfy. Secondly, we define distortion, spectral, entropic, generalized quantiles and Haezendonck-Goovaerts risk measures by construction. Thirdly, there are metrics based on statistical properties, such as those defined based on moments. One risk measure can belong to multiple categories. For example, Tail Value-at-Risk (TVaR) is a spectral risk measure along with being coherent and a distortion risk measure. Moreover, risk measures are utilized both for single random variables and multivariate random variables. The multivariate random variables contain risk measures defined with dependence structures among random variables. Also, there are parametric risk measures, semi-parametric risk measures and non-parametric risk measures, given the experienced data or the parameters for certain specially distributed random variables are sufficiently provided.

Definition Let $X$ be a random variable such that the risk measure of $X$, $\rho(X)$, is a functional with $\rho : X \to (-\infty, \infty]$ with $\rho(\mathcal{L}^\infty) \subset \mathbb{R}$. In actuarial science, we define a risk measure $\rho : \chi \to (-\infty, \infty)$ as mapping of a random variable from the probability space to the real line: $\mathbb{R}$. Risk measures are crucial for quantifying risks and translating bulks of data into easy-to-understand real numbers. For example, the expectation $E(X)$ of a random variable $X$ is a risk measure because it gives us a estimation so we can grasp a feeling of the uncertainty of the risk.

Risk measures are denoted as $\rho(X)$ in this paper. We can think of $\rho(X)$ as a function that we use to derive a monetary amount to prevent the loss that may be causeds by $X$. In life insurance companies, $\rho(X)$ is paramount because the company will assign the premiums they collect to the reserve based on their estimation of $\rho(X)$. If the monetary value of their reserves are insufficient to cover the losses incurred, the company may go into bankruptcy.

2.1 Risk Measures by Axioms

Coherent Risk Measures A majority of the risk measures that we will discuss are coherent. Artzner et al. (1999) defined that a risk measure is coherent if it satisfies the following four properties: monotonicity, positive homogeneity, sub-additivity and translation invariance.
The definitions of the axioms are as follows:

- Monotonicity: \( \rho(X) \leq \rho(Y) \), if \( X \leq Y \).
- Positive Homogeneity: \( \rho(\lambda X) = \lambda \rho(X) \), \( \lambda \in (0, \infty) \).
- Subadditivity: \( \rho(X + Y) \leq \rho(X) + \rho(Y) \).
- Translation Invariance: \( \rho(X + c) = \rho(X) + c \), \( c \in \mathbb{R} \).

**Convex Risk Measures**  Convex risk measures are used as a way to introduce a better diversification benefit compared to coherent risk measures. The idea of convexity is that it takes the properties of positive homogeneity and subadditivity, and combines them to better portray the liquidity risk of a portfolio. Due to this combination of other properties, convexity is a weaker property than positive homogeneity and subadditivity.

Convexity:

\[ \rho(\lambda X + (1-\lambda)Y) \leq \lambda \rho(X) + (1-\lambda)\rho(Y) \text{ for any } \lambda \in [0, 1]. \]

A map \( \rho : X \to \mathbb{R} \) of a convex risk measure must also satisfy the properties of monotonicity and translation invariance to be valid.

### 2.2 Risk Measures by Construction

#### 2.2.1 Distortion Risk Measures

Distortion risk measures are first introduced in Yaari (1987). A distortion function \( g : [0, 1] \to [0, 1] \) is a non-decreasing function with \( g(0) = 0 \) and \( g(1) = 1 \). For a random loss variable \( X \) with decumulative distribution function \( S(X) = 1 - F(X) \), we have the distortion risk measure:

\[ \rho_g(X) = \int_0^\infty g(S(x)) \, dx. \]

As the name implies, the distortion function adjusts the true probability of events by giving more weight to higher risk events.

- Value at Risk

**Definition**  Value at Risk (VaR) is the amount of losses at a given confidence level \( \alpha \). A definition of VaR given by Linsmeier and Pearson (2000), with a probability of \( x \) percent and a holding period of \( t \) days, an entity’s VaR is the loss that is expected to
be exceeded with a probability of only $x$ percent during the next $t$-day holding period. Alternatively, VaR is the loss that is expected to be exceeded during $x$ percent of $t$-day holding periods. VaR is the most widely used risk measure as it is an easy value to calculate from historical data or probability distributions.

For $p \in (0, 1)$ and random variable $X$, $F_X(x)$ is the corresponding cumulative distribution function, VaR is defined as

$$VaR_p(X) = F_X^{-1}(p) = \inf \{ x \in \mathbb{R} : F_X(x) \geq p \}.$$ 

**Properties** VaR follows the properties of monotonicity, positive homogeneity and translation invariance. It is also considered a monetary risk measure because it satisfies monotonicity and translation invariance. The big drawback to using VaR over other risk measures is VaR is not coherent. It fails the property of subaddativity, as the VaR of a combined portfolio can be larger than the individual portfolios.

**Distortion Function** As stated by Hardy and Wirch (2002), the distortion function for VaR is defined as

$$g(t) = \begin{cases} 
1, & \text{if } 1 - \alpha < t \leq 1, \\
0, & \text{if } 0 < t < 1 - \alpha. 
\end{cases}$$

• **Expected Shortfall**

**Definition** Expected Shortfall (ES) is sometimes known as Tail Value at Risk (TVaR). Depending on the source, ES also goes by a variety of other names. The most common are Conditional Value at Risk (CVaR), Average Value at Risk (AVaR), and expected tail loss (ETL). All of these other risk measures produce the same value as ES when using the same data.

The ES of a random variable $X$ at the $\alpha$-th probability level $ES_\alpha[X]$ is measured by the ‘arithmetic average’ of VaR of X from $\alpha$ to 1:

$$ES_\alpha[X] = \frac{1}{1 - \alpha} \int_\alpha^1 VaR_p[X] dp.$$ 

Generally, we have the basic identity that proves ES is not less than VaR. That identity is
\[ ES_\alpha[X] = VaR_\alpha[X] + \frac{1}{1-\alpha} \mathbb{E}[\max(X - VaR_\alpha(X), 0)]. \]

For a random variable \( X \) with continuous distribution, we have the conditional expected loss given that the loss exceed \( VaR_\alpha(X) \) such as

\[ ES_\alpha[X] = \mathbb{E}[X|X \geq VaR_\alpha(X)] = \mathbb{E}[X|X > VaR_\alpha(X)]. \]

Thus, the relationship between ES and VaR has identified that ES as the minimum cost function defined by the model that is valid for any type of distribution

\[ ES_\alpha[X] = \min_p \frac{1}{1-\alpha} \mathbb{E}[\max(X - p, 0)] + p, \]

where the smallest \( p \) solving the minimization equation is \( VaR_\alpha(X) \).

**Properties**  One of the advantages of ES is its property of coherence, meaning that it satisfies subaddativity, monotonicity, translation invariance and positive homogeneity. ES is also comonotonic additive. A disadvantage to using ES over VaR, however, is the lack of elicitability in most cases, which makes it very difficult to backtest. ES provides a more conservative value compared to VaR at the same confidence level as it takes the average of the data in the tail at confidence level \( \alpha \) rather than the value at \( \alpha \).

**Distortion Function** The distortion function for ES is defined as

\[ g(t) = \begin{cases} 
1 & \text{if } 1-\alpha \leq t \leq 1, \\
\frac{t}{1-\alpha} & \text{if } 0 \leq t < 1-\alpha.
\end{cases} \]

- Conditional Tail Expectation (CTE)

**Definition** Brazauskas et al. (2008) defined CTE as: given a loss variable \( X \) (which is a real-valued random variable) with finite mean \( \mathbb{E}[X] \), and let \( F_X \) denote its distribution function. Next, let \( F_X^{-1} \) be the left-continuous inverse of \( F_X \). For every \( t \in [0, 1] \), we have

\[ F_X^{-1}(t) = \inf\{x : F_X(x) \geq t\}. \]
With these notations, CTE is defined by

\[ CTE_t(X) = \mathbb{E}[X | X > F_X^{-1}(t)]. \]

If \( F_X \) is continuous, then \( F_X(F_X^{-1}(t)) = t \) for every \( t \in [0,1] \). In other words, at point \( t \), \( t \cdot 100\% \) of losses are at or below \( t \), while \( (1-t)100\% \) of losses are above \( t \). In the continuous case, CTE is defined as

\[ CTE_t(X) = \frac{1}{1-t} \int_t^1 F_X^{-1}(u)du. \]

**Properties** The main property that CTE has is that it is a coherent risk measure only in the continuous case.

**Relation to ES** As noted by their definitions, ES and CTE are similar risk measures. ES and CTE are equal to each other if the distribution is continuous and calculated at the same value of \( \alpha \), otherwise they might be different. This result can be expected as both ES and CTE calculate the expected loss in the right tail of the distribution. In the continuous case, there is only one value for each distribution at \( \alpha \).

### 2.2.2 Spectral Risk Measures

**Definition** Spectral risk measures involve a weighted average of the quantiles of a loss as stated by Adams et al. (2007). A spectral risk measure is defined as

\[ M_\phi(X) = - \int_0^1 \phi(p) F_X^{-1}(p) dp. \]

The function \( \phi \) is right-continuous, non-negative and non-increasing. It is defined from \([0,1] \) and \( \int_0^1 \phi(p) dp = 1 \). As mentioned in Acerbi (2002), an admissible risk spectrum \( \phi \in \mathcal{L}(\mathbb{R}, [0,1]) \) will be called the ‘risk aversion function’ of the risk measure \( M_\phi(X) \), where \( M_\phi \) will be called the ‘spectral risk measure’.

**Properties** Spectral risk measures satisfy monotonicity, positive homogeneity and translation invariance, but also include other properties that make them robust than other risk measures. Those properties are law-invariance and comonotonicity as defined below:

Law-Invariance: For \( X \) and \( Y \) with cumulative distribution functions \( F_X \) and \( F_Y \), if \( F_X = F_Y \),
then $\rho(X) = \rho(Y)$;
Comonotonicity: $\rho(X + Y) = \rho(X) + \rho(Y)$ for every comonotonic random variables $X$ and $Y$.

In order to determine if $X$ and $Y$ are comonotone to each other, the following property must hold:

For every $\omega_1, \omega_2 \in \Omega : (X(\omega_2) - X(\omega_1))(Y(\omega_2) - Y(\omega_1)) \geq 0$.

**Expected Shortfall** Expected Shortfall is the main example of a spectral risk measure. This measure, as referenced to in Adam et al. (2007), is ‘any coherent risk measure defined on a finite space $\Omega$ can be set under the following form

$$\rho(X) = -\inf\{\mathbb{E}^Q[X] | Q \in \Pi\},$$

where $\Pi$ is a set of probability measures on $Q$ defined on $\Omega$.’ Using this definition, we obtain a definition for ES at confidence level $\alpha$ to be defined as

$$ES_\alpha(X) = -\inf\{\mathbb{E}^Q[X] | Q \in \Pi_\alpha\}.$$  

The dataset $\Pi_\alpha$ consists of the data points above $\alpha$ and applying a constant weight to those points, while assigning a value of 0 to every data point below $\alpha$. Expected Shortfall is a spectral risk measure with the risk aversion function $\phi(p) = \frac{1}{\alpha} * 1_{[0,\alpha]}(p)$.

### 2.2.3 Entropic Risk Measures

According to Follmer and Knispel (2011), an Entropic Risk Measure is modeled by:

$$e_\gamma(Z) := \rho(Z) = \frac{1}{\gamma} \log \mathbb{E}_P[e^{-\gamma X}] = \sup_Q \{\mathbb{E}_Q[-X] - \frac{1}{\gamma} H(Q|P)\}$$

for parameters $\gamma \in [0, \infty)$, where $e_0(X) := \mathbb{E}_P[-X]$ and $H(Q|P)$ with definition:

$$H(Q|P) = \begin{cases} 
\mathbb{E}_Q[\log \frac{dQ}{dP}] & \text{if } Q \ll P, \\
+\infty & \text{otherwise.}
\end{cases}$$

and denotes the relative entropy of $Q$ with respect to $P$. In addition, it is shown that the model is increasing in $\gamma$ (Follmer and Knispel, 2011). Furthermore, Yongchao and Huifu(2014) found that the model converges to the essential supremum given optimum condition while holding such properties; smooth distributions and significantly weighs on tail when $\gamma$ is large enough. Meanwhile, Yan (2015) studied the deviation for parameter and level variations
related to its convex and coherent model; it is identified that $\text{ENT}_p^\gamma(X)$ is the convex entropic risk measure with parameter $\gamma$ as the risk aversion parameter. Let $X$ be a random variable on probability space $(\Omega, F, P)$ with $\text{ENT}_p^\gamma(X)$ is defined as

$$\text{ENT}_p^\gamma(X) := \frac{1}{\gamma} \log \mathbb{E}_p(e^{-\gamma X}),$$

where $\mathbb{E}_p(\cdot)$ means the mathematical expectation with respect to $P$ and $\gamma > 0$.

**Properties of $\text{ENT}(\gamma, p)$** As implied above, $\text{ENT}_p^\gamma(X)$ satisfies the convexity property for risk measures, but it is not a coherent risk measure as it does not satisfy the property of positive homogeneity.

$\text{CERM}_p^c(X)$ is the coherent entropic risk measure model with parameter $c$ as the level. For each $c > 0$, $\text{CERM}_p^c(X)$ is defined as

$$\rho_c(X) := \sup_{\{Q \in M_1 : H(Q|P) \leq c\}} \mathbb{E}_Q[-X], \quad X \in L^\infty,$$

where $M_1$ denotes the class of all probability measures on $\mathcal{X}$ with a functional $\rho : \mathcal{X} \in \mathbb{R}$ with properties monotonicity and translation invariance. Another, more concise definition for $\text{CERM}_p^c(X)$ is

$$\text{CERM}_p^c(X) = \inf_{\gamma > 0} \left\{ \frac{c}{\gamma} + \text{ENT}_p^\gamma(X) \right\}$$

**Properties of $\text{CERM}_p^c(X)$** $\text{CERM}_p^c(X)$ satisfies the property of coherence and is also law-invariant.

**Entropic Value at Risk** Ahmadi-Javid (2012) introduced Entropic Value-at-Risk that shows the corresponding *tightest possible upper bound* derived from the Chernoff Inequality. In this case, the Chernoff Inequality for any constant $a$ and $X \in L_{M+}$ is defined as

$$\Pr(X \geq a) \leq e^{-za}M_X(z), \quad \forall z > 0.$$  

When solving the equation $e^{-za}M_X(z) = a$ with respect to $a$ for $\alpha \in [0, 1]$, the following equation is obtained:

$$a_X(\alpha, z) := z^{-1}\ln\left(\frac{M_X(z)}{\alpha}\right).$$

This is proven as one of the coherent risk measure defined by the model:
\( \text{EVAR}_{1-\alpha}(X) := \inf_{z > 0} \{a_X(\alpha, z)\} = \inf_{z > 0} \{z^{-1} \ln \left( \frac{M_X(z)}{\alpha} \right) \}. \)

This definition had leading us to a proposition that this risk measure also depends on the moment generating function. He also showed the dual representation through Donsker-Varadhan Variational Formula which is

\[
\text{EVAR}_{1-\alpha}(X) = \sup_{Q \in \mathcal{Z}} \mathbb{E}_Q(X),
\]

where \( \mathcal{Z} = \{Q \ll P : D_{KL}(Q||P) \leq -\ln \alpha \} \) and \( D_{KL}(Q||P) := \int \frac{dQ}{dP} (\ln \frac{dQ}{dP}) dP \) is the relative entropy of \( Q \) with respect to \( P \). Furthermore, he demonstrated that \( \text{EVar} \) is the upper bound of for both \( \text{VaR} \) and \( \text{CVaR} \) at the same level of confidence which means that \( \text{EVar} \) is known to be as more risk averse compared to others. Through financial view, \( \text{EVar} \) requires a lot of resources allocation for least possible risk yet made it undesirable to be used.

**Properties**  Unlike other risk measures that involve \( \text{VaR} \), \( \text{EVaR} \) is a coherent risk measure.

### 2.2.4 Generalized Quantiles

It is important to know the properties that the generalized quantiles of a random variable \( X \) refers to the ‘minimizers of a piecewise-linear loss function’ (Bellini et al., 2014). The metrics are defined as

\[
q_\alpha(X) = \arg \min_{x \in \mathbb{R}} \{\pi_\alpha(X, x)\},
\]

where \( \pi_\alpha(X, x) = \alpha \mathbb{E}[(X - x)^+] + (1 - \alpha) \mathbb{E}[(X - x)^-] \) with \( x^+ = \max\{x, 0\} \) and \( x^- = \max\{-x, 0\} \). Furthermore, we can identify generalized quantile as first-order condition as any minimizer such that \( x_\alpha^* \in \arg \min \{\pi_\alpha(X, x)\} \) is proven to be a generalized quantile. Here, \( \pi_\alpha(X, x) \) satisfies the properties of finite, non-negative and convex in a closed interval. According to Bellini et al. (2014), there are numerous generalized quantile risk measures in the literature such as expectiles, power loss functions, Orlicz quantiles, generic loss functions all holding similar properties: translation invariance, constancy, internality, monotonicity, positive homogeneity and convexity. However, Orlicz quantiles lack the property of monotonicity. In addition, only expectiles are a type of coherent generalized quantile due to strict monotonicity. This type of expectiles can be written as

\[
e_\alpha(X) - \mathbb{E}[X] = \frac{2\alpha - 1}{1 - \alpha} \mathbb{E}[(X - e_\alpha(X))^+].
\]
Bellini et al. (2014) also added that expectiles are a more conservative measure than other heavy tailed quantile distributions for large \( \alpha \).

### 2.2.5 Haezendonck-Goovaerts (HG) Risk Measure

This risk measure introduced by Haezendonck and Goovaerts (1982) is based on the zero utility premium principle as normalized Young functions and Orlicz Norms. The mean value principle premium calculation is a rule that assigns a number \( H(X) \) to any given risk \( X \) as defined by

\[
v(H(X)) = \mathbb{E}[v(X)],
\]

where such \( v \) is chosen that \( v' > 0 \) and \( v'' \geq 0 \). If a reinsurer takes on part of this risk. A Young function is defined as a mapping from \( R_0^+ \) into \( R_0^+ \) with integral form

\[
\Phi(x) = \int_0^x g(t) dt,
\]

where \( g(t) \) is a left-continuous, monotone increasing real-valued function on \( R_0^+ \) with \( g(0) = 0 \) and \( \lim_{x \to \infty} g(x) = \infty \). The function \( g \) is called the kernel of the Young function \( \Phi \). A Young function is said to be normalized if \( \Phi(1) = 1 \). If \( \Phi(1) > 0 \), the Young function can be normalized by taking \( \frac{\Phi(x)}{\Phi(1)} \). Given \( \Phi \) is a Young function, and if \( X \in L_1^+ \) where \( L_1^+ \) is defined as

\[
L_1^+ = \{X \in L_1 | X \geq 0 \text{ a.s.}\}
\]

with \( X \neq 0 \) a.s., then

\[
\Psi(x) = \mathbb{E}[\Phi(X/x)]
\]

is a mapping from \( \mathbb{R}^+ \) into \( R_0^+ \cup \infty \). \( \Psi(x) \) is also a Young function.

**Properties of \( \Psi(x) \)**

- \( \Psi \) is right continuous at every \( x \in \mathbb{R}^+ \) and continuous at every interior point of \( (\Psi < \infty) \),
- \( \Psi \) is monotone decreasing on \( \mathbb{R}^+ \) and is (strictly) decreasing on \( (\Psi < \infty) \),
- \( \lim_{x \to 0} \Psi(x) = \infty \),
- \( \lim_{x \to \infty} \Psi(x) = 0 \) if \( (\Psi < \infty) \neq \emptyset \).

The next major subtopic for HG risk measures are Orlicz spaces and Orlicz norms. The Orlicz space is defined as the set \( L_\Phi \) of random variables \( X \) such that:
\[ \mathbb{E}[\Phi(|X|/a)] \leq 1 \quad \text{for some } a > 0, \]

which is a subspace of \( L_1 \). The Orlicz norm on the Orlicz space \( L_\Phi \) is defined to be:

\[ ||X||_\Phi = \inf\{a > 0 | \mathbb{E}[\Phi(|X|/a)] \leq 1 \}. \]

The Orlicz norm follows the properties of positive homogeneity and sub-additivity. The norm is always greater than 0 and equals 0 if and only if \( X = 0 \). Using the formulas of the Young functions we can calculate the HG premium principle for bounded risks. If \( X \in L_\Phi^+ \) and \( X \neq 0 \) a.s., then the equation

\[ \Psi(x) = \mathbb{E}[\Phi(X/x)] = 1 \]

has exactly one solution denoted by \( H(X) \) with \( H(X) = 0 \) if \( X = 0 \) a.s. and \( H(X) \) is named as the H-G risk measure for \( X \).

**Properties of \( H(X) \)**

- \( H(X) = H(Y) \) if \( F_X = F_Y \),
- if \( X \) is a constant \( K \) a.s. then \( H(X) = K \),
- \( \mathbb{E}[X] \leq H(X) \),
- \( H(X + Y) \leq H(X) + H(Y) \),
- \( H(aX) = aH(X) \) if \( a \in \mathbb{R}_0^+ \),
- \( H(X) \leq H(Y) \) if \( X < Y \),
- if \( X \leq \) constant \( K \) a.s., then \( H(X) \leq K \),
- if \( \Phi \) is strictly convex, then \( \mathbb{E}[X] < H[X] \) except when \( X = \) constant a.s.

Other than being homogenous and translation invariant, it is found that HG risk measure, with \( \phi \) derived from a concave distortion function \( g \) is sub-additive (Goovaerts et al., 2012). Also, the HG risk measure is an application of the mean value principle. Additionally, according to Bellini and Gianin (2011), HG risk measures are not comonotonically additive yet the simplest coherent risk measures. This type of risk measure is naturally defined on Orlicz spaces. HG premium is identified to be finite, convex, law-invariant and coherent (Bellini and Gianin, 2011).
2.3 Statistical Risk Measures

**Expectation-first moment**  Expectation of a random variable is defined by $\mathbb{E}[X]$, which is the simplest risk measure to evaluate the average loss.

**Variance and Standard Deviation-second moment**  Variance is usually adopted to describe the deviation from mean, and is defined by

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

for a random variable $X$. Moreover, the most commonly used deviation risk measure is the standard deviation $\sigma$, which is calculated as $\sigma(X) = \sqrt{Var(X)}$. Deviation measures are used to determine how far from the mean of the data the data is distributed. Deviation is critical in the world of finance as having the ability to predict what future gains or losses one may have can lead to better management of wealth. In general, higher standard deviation means more risk as the spread of potential returns is spread out more from the mean. By Rockafellar et al. (2002), a deviation measure on $\mathcal{L}^2$ will mean any functional $D: \mathcal{L}^2 \to [0, \infty]$ satisfying the following five properties: shift invariant, normalization, positive homogeneous, subadditivity and positivity. The additional properties are defined as below,

- Normalization: $D(0) = 0$.
- Positivity: $D(X) > 0$ for nonconstant $X$ and $D(X) = 0$ for constant $X$.
- Shift-Invariance: $D(X + c) = D(X)$ for $c \in \mathbb{R}$.

**Skewness-third moment**  The skewness of a distribution is a measure of the asymmetry of a probability distribution about its mean. Skewness is the third standardized moment of a distribution, and is calculated by

$$\gamma_1 = \mathbb{E}[(\frac{X - \mu}{\sigma})^3] = \frac{\mu_3}{\sigma^3},$$

where $\mu_3$ is the third centralized moment of a distribution. A skewness value of 0 means that the distribution is perfectly symmetrical about its mean. A skewness value below 0 indicates that the distribution has a “longer left tail”. In other words, more data lies farther out in the left tail while a majority of the data is bunched up on the right side of the distribution. For a skewness value above 0, the opposite is true. The right tail of the distribution has more data farther away from the mean while a majority of the data is bunched up on the left side of the distribution. When assessing risk measures, the skewness value can be used to
help determine how far away from the mean the confidence level ($\alpha$) value will be. A higher absolute value of skewness shows that there are outliers in the data set, which can affect the difference in value between VaR and TVaR greatly.

**Kurtosis-fourth moment** The kurtosis of a distribution describes the shape of a curve in terms of how the data is plotted throughout a probability distribution. Like skewness, kurtosis is also a standardized moment with it being the fourth standardized moment and is calculated by

$$
\gamma_2 = \mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4}.
$$

Kurtosis determines how "narrow and tall" or "wide and short" the distribution is. In other words, kurtosis measures the tailed-ness of a distribution. Kurtosis is a positive value with higher values representing greater tailed-ness. Kurtosis values for distributions are often compared to the value for a standard normal distribution to better understand the shape of the distribution. The kurtosis value for a standard normal distribution is 3. A value less than this means the data is highly centered with few outliers, and therefore small tails. A value greater than 3 means the distribution has lots of extreme outliers and data far from the mean, resulting in heavy tails.

**Semi-variance** Semi-variance, like the name implies, does not compute the variance of the whole data set that is being used. Instead, semi-variance only is the variance for a portion of a data set. In finance and portfolio selection, the semi-variance of a data set is typically calculated using only the data points below the mean or target return of that data set. Semi-variance can be used by investors to determine how much downside risk they are taking on. For a random variable $X$, the semi-variance is defined by

$$
Var^+(X) = \mathbb{E}\left[(X - \rho[X])^2|X > \rho[X]\right],
$$

where $\rho$ is the risk measure to denote the target of the risk.

**Tail-variance** Like semi-variance, tail-variance (TV) does not calculate the variance of the whole distribution, but only a part of it. In this case, tail-variance only calculates the variance of data points located in the tail of the distribution, where the tail is defined above a confidence level $\alpha$. Furman and Landsman (2006) defined tail-variance to be

$$
TV_{\alpha}(X) = Var(X|X > x_\alpha) = \mathbb{E}\left((X - \tau_{\alpha}(X))^2|X > x_\alpha\right),
$$
where $\tau_\alpha(X) = TCE_\alpha(X) = \mathbb{E}(X|X > x_\alpha)$. TV is used to measure how much risk is located within the tail itself, and to determine how spread out the data points are.

**Region-variance** Tail-variance as Furman and Landsman (2006) defined can be generalized. For a random variable $X$, the region-variance (RV) can be defined by

$$RV = Var(X|A) = \mathbb{E}[(X - RE_A(X))^2|X \in A],$$

where $RE_A$ is the region-expectation defined by $RE_A(X) = \mathbb{E}[XI_A]$ and $A$ is the special region considered.

### 2.4 Dependence Risk Measures

#### 2.4.1 CoVariance

Covariance is defined by

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])],$$

for two random variables $X$ and $Y$ to illustrate their linear dependence structure.

#### 2.4.2 Dependence Risk Measure by comparison with Comonotonicity and Independence (Dhaene et al., 2014)

Let $\mathbf{X}^c = (X_1^c, \ldots, X_d^c)$ be a random vector with the same marginal distributions as $\mathbf{X}$ but with comonotonic components and $S^c = \sum_{i=1}^d X_i^c$. In addition, $\mathbf{X}^\perp = (X_1^\perp, \ldots, X_d^\perp)$ be a random vector with the same marginal distributions as $\mathbf{X}$ but with comonotonic components and $S^\perp = \sum_{i=1}^d X_i^\perp$, Dhaene et al. (2014) defined

$$\rho_c(\mathbf{X}) = \frac{Var(S) - Var(S^\perp)}{Var(S^c) - Var(S^\perp)} = \frac{\sum_{i=1}^d \sum_{j<i} Cov(X_i, X_j)}{\sum_{i=1}^d \sum_{j<i} Cov(X_i^c, X_j^c)}$$

provided the covariances exist. This risk measure satisfies normalization, monotonicity, permutation invariance and duality.
3 Robustness

3.1 Introduction

Generally speaking, robustness is used with risk measures to determine how well they stand up to small or large changes in the underlying datasets. A robust model is one that unaffected by outliers or small errors from the assumptions made in the model.

Developing a robust risk measure is of utmost importance to corporations, especially financial institutions, as they use hundreds of different models and distributions. Applying a risk measure that is not robust can lead to issues as small errors or deviations in the assumptions or data could drastically affect the output data. In the aftermath of the financial crisis, people start to realize that the robustness of the estimate is important. Consequently, regulators and other stakeholders have started to require that the internal models used by financial institutions are robust.

In the following discussion, we are going to exhibit different measurements of robustness in risk measure process and their characteristics. Also, we will analysis traditional risk measure method like VaR and TVaR in those robustness measurements.

3.2 Qualitative Robustness

Informally, qualitative robustness refers to a certain insensitivity of the sampling distribution with respect to deviations from the theoretical distribution. We will focus on law invariant risk estimators in this part. A risk estimator is said to be robust if a small variation in the loss distribution results in a small change in the the risk estimation.

3.2.1 Definition

Definition 3.1. (Cont, 2010) A risk estimator $\hat{\rho}$ is qualitative robust at $F$ if for any $\epsilon > 0$ there exist $\delta > 0$ and $n_0 \geq 1$ such that

$$d(F, G) \leq \delta \Rightarrow d(\hat{\rho}(F), \hat{\rho}(G)) \leq \epsilon, \forall n \geq n_0,$$

where $C$ is a fixed set of loss distributions and $F \in C$.

The intuitive notion of robustness in this paper can now be made more precisely by adopting this definition. With respect to the above definition, $d(F, G) \leq \delta$ indicates that the distortion
level of distribution is bounded in certain radius $\delta$, meaning the variation is so small that the value from risk function will only make a small change less than $\epsilon$. However, Definition 3.1 is not widely used in either econometric or financial applications since it cannot give us a quantitative result of robustness. Moreover, the following characteristics (a) and (b) will also limit its application in practice:

(a) It may not be able to solve different behaviors in tail distribution.

In practice, two distributions can be rather close with respect to a distance $d$, but still have completely different tail behavior. In this case, whether the risk functional is sensitive to tail behavior is determined by the metrics selected. Here we provide one example for each case.

**Example 1: Lévy metric**

**Definition 3.2.** (Kratschmer, 2012) Let $F_\mu$ and $F_\nu$ be the distribution functions for parameters $\mu$ and $\nu$. Then, Lévy metric between these two distributions can be defined as follow:

$$d_{\text{Levy}}(F_\mu, F_\nu) = \inf \{ \epsilon > 0 : F_\mu(x - \epsilon) - \epsilon \leq F_\nu(x) \leq F_\mu(x + \epsilon) + \epsilon \text{ for all } x \}.$$  

Intuitively, if between the graphs of $F_\mu$ and $F_\nu$ one inscribes squares with sides parallel to the coordinate axes, then the side-length of the largest such square is equal to $d_{\text{Levy}}(F_\mu, F_\nu)$.

To illustrate the point that Lévy metric is insensitive in tail distribution, let $\mathcal{M}_1$ be the class of all probability measures on $\mathbb{R}$ and recall that the Lévy metric metrizes the weak topology on $\mathcal{M}_1$ and that the compactly supported probability measures are dense in $\mathcal{M}_1$ with respect to weak convergence. Hence, for every $\mu \in \mathcal{M}_1$ and every $\epsilon > 0$ there exists a compactly supported $\nu \in \mathcal{M}_1$ such that $d_{\text{Levy}}(F_\mu, F_\nu) < \epsilon$. Clearly, $F_\mu$ can have arbitrary tail behavior, whereas the tail behavior of $F_\nu$ is trivial.

**Example 2: Wasserstein and $L^p$ metric**

To illustrate that Wasserstein metric is sensitive to tail distribution, we can have a simple simulation example to compare Wasserstein metric and $L^p$ distance. The definition of Wasserstein distance is defined by Definition 3.3 and $L^p$ metric by Definition 3.4 as below.

**Definition 3.3.** (Kiesel, 2016) Let $F, G$ be two distribution functions. The Wasserstein distance, $W_p(F, G)$ for two distributions is given by

$$W_p(F, G) = \int_0^1 |F^{-1}(u) - G^{-1}(u)|^p du, \quad p \geq 1.$$
Table 1: Wasserstein and $L^p$ distance, $\lambda=0.1$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\epsilon$</th>
<th>$L^p$ Distance</th>
<th>Wasserstein Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>p=1</td>
<td>0.05</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td>0.075</td>
<td>0.800</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>p=2</td>
<td>0.05</td>
<td>0.106</td>
<td>0.455</td>
</tr>
<tr>
<td></td>
<td>0.075</td>
<td>0.169</td>
<td>1.164</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.213</td>
<td>1.820</td>
</tr>
<tr>
<td>p=3</td>
<td>0.05</td>
<td>0.069</td>
<td>0.632</td>
</tr>
<tr>
<td></td>
<td>0.075</td>
<td>0.109</td>
<td>2.585</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.137</td>
<td>5.053</td>
</tr>
<tr>
<td>p=4</td>
<td>0.05</td>
<td>0.057</td>
<td>1.164</td>
</tr>
<tr>
<td></td>
<td>0.075</td>
<td>0.091</td>
<td>7.625</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.114</td>
<td>18.634</td>
</tr>
</tbody>
</table>

**Definition 3.4.** (Kiesel, 2016) $L^p$-distance, $\Theta_p(F, G)$ is defined as

$$\Theta_p(F, G) = \left( \int_{-\infty}^{\infty} |F(u) - G(u)|^p du \right)^{1/p}, \quad p \geq 1.$$  

We compare $L^p$ distance with the Wasserstein metric for a standard normal distribution $\Phi$, by which we perturbate data such that there is a difference in the right tail. Thus, let

$$G(x) = (1 - \epsilon)\Phi(x) + \epsilon Exp(x, \lambda),$$

where $Exp(x, \lambda)$ is an exponential distribution with parameter $\lambda$.

**Example 3.1.** Suppose $\lambda=0.1$, we get the values of $L^p$ distance and Wasserstein distance in Table 1 by changing values of $p$ from 1 to 4. As is illustrated by Figure 1 and 2, the Wasserstein distance increases in $p$ while the $L^p$ distance decreases. Thus, the weighting of differences in the tail increase for the Wasserstein metric in $p$, while it decreases for $L^p$.

In the recent years of financial crisis, it has become apparent that a misspecification of the tail behavior of loss can largely influence the result of risk measure. As a result, applying qualitative robustness measure without choosing specific risk functional or metric could lead to a dramatic underestimation of the associated risk.
Figure 1: $L^p$ distance as to $p$

Figure 2: Wasserstein distance as to $p$
(b) It is clustering all risk functional into two groups.

The qualitative robustness definition generates a sharp division of risk functionals into the class of those that are called ‘robust’ and another class of those that are called ‘not robust’. However, the distinction between ‘robust’ and ‘non-robust’ risk functionals is artificial because there is actually a full continuum of possible degrees of robustness beyond the classical concept. So labeling a risk measure as ‘robust’ or ‘non-robust’ may give a false impression.

In addition, we will introduce Theorem 3.1 which provides an equivalence between the loss distribution set and qualitative robust for a risk measure.

**Theorem 3.1.** (Cont, 2010) Let $\rho$ be a risk measure, $C$ is a fixed set of loss distributions and $F \in C$. If $\hat{\rho}^h$ is consistent with $\rho$ at every $G \in C$, the following are equivalent:

(a) the restriction of $\rho$ to $C$ is weakly continuous at $F$;
(b) $\hat{\rho}^h$ is qualitative-robust at $F$.

**VaR** For any function $F \in C$ where $C$ is the set of all distributions continuous at $\alpha$ if

$$q_F^+(\alpha) = q_F^-(\alpha), \alpha \in (0, 1).$$

Then, historical $\text{VaR}_\alpha$ is qualitative robust at $F$, and hence, $\text{VaR}_\alpha$ is weakly continuous at $F$ by Cont (2010). By adopting Theorem 3.1, we know that $\text{VaR}_\alpha$ is qualitative robust.

**TVaR** The statistical functional of $\text{TVaR}$ is

$$T(F) = -(1 - \alpha)^{-1} \int_{-\alpha}^1 F^{-1}(x)dx.$$  

Historcial TVaR function is not weakly continuous at $D = \{F \in \mathcal{P}, \int_{-\infty}^{+\infty} |x|dF < \infty\}$ according to Cont (2010), and thus not qualitative robust by Theorem 3.1. However, TVaR is robust by Wasserste distance, which has been proved by Kiesel (2016). In summary, TVaR is generally not robust unless certain distribution distance metric is applied. Note that fully proof that Corollary needs other definitions and Lemma from additional references. For simplicity, we only include the conclusion that TVaR is not weakly continuous on any distributions with finite first moment.
3.3 Sensitivity analysis

3.3.1 Definition and Property

**Definition 3.5.** (Cont, 2010) The sensitivity function of a risk measurement procedure is defined by

\[
S(z, F) = \lim_{\epsilon \to 0^+} \frac{\rho(\epsilon \delta_z + (1 - \epsilon)F) - \rho(F)}{\epsilon}
\]

for any \( z \in \mathbb{R} \) such that the limit exists.

Here, \( S(z, F) \) measures the sensitivity of the risk estimator based on a large sample to the addition of a new data point. The sensitivity function measures the directional derivative of the risk measure function \( \rho \) at \( F \), and \( \epsilon \) in the function denotes the distorted level of distribution.

Sensitivity analysis is a numeric measurement of robustness. Unlike the definition given by qualitative robust, sensitivity analysis can give us a quantitative result to show the degree of robustness. If the risk function has a bounded sensitivity, it means that the risk measurement procedure is not quite sensitive to a small change in the data set. On the contrary, if the risk function has an unbounded sensitivity, it means that this risk measurement procedure is relatively less robust.

3.3.2 Sensitivity analysis for VaR and TVaR

As mentioned above, sensitivity of a risk estimator is defined as the directional derivative of distorted level \( \epsilon \). Based on the prior discussion, VaR should be robust than TVaR, meaning that the sensitivity of VaR should be bounded while TVaR with a unbounded sensitivity function.

According to Cont (2010), we will analyze the sensitivity testing for VaR and TVaR as below.

**VaR**  The sensitivity function at \( F \) of the historical \( VaR_\alpha \) is

\[
S(z) = \begin{cases} 
\frac{1 - \alpha}{f(q_\alpha(F))}, & \text{if } z < q_\alpha(F), \\
0, & \text{if } z = q_\alpha(F), \\
-\frac{\alpha}{f(q_\alpha(F))}, & \text{if } z > q_\alpha(F).
\end{cases}
\]
Figure 3: VaR and TVaR with 100 replications per day

**TVaR** The sensitivity function at $F \in \mathcal{D}^-$ for historical $TVaR_\alpha$ is

$$S(z) = \begin{cases} 
-\frac{z}{\alpha} + \frac{1-\alpha}{\alpha} q_\alpha(F) - TVaR_\alpha(F), & \text{if } z \leq q_\alpha(F), \\
-q_\alpha(F) - TVaR_\alpha(F), & \text{if } z \geq q_\alpha(F).
\end{cases}$$

From these risk functions, we can see that historical VaR has a bounded sensitivity whilst TVaR has a linear sensitivity, indicating that VaR is robuster than TVaR.

### 3.3.3 Numerical Example

As shown above, VaR is a robuster risk functional than TVaR. To start the simulation, we propose an empirical investigation to show the above conclusions. More precisely, we use historical simulation to generate random samples from a dataset of S&P 500.

Figure 3 plots the values of the historical VaR and TVaR for a total of 100 replications per day. We can see how the overall path of historical VaR is more regular than that of TVaR, which is more volatile. Thus, we can conclude that the historical estimator of TVaR is less robust, confirming the insights we get above.

Furthermore, we can see how VaR and TVaR change with some small perturbation to the loss.
Figure 4: VaR and TVaR by adding perturbation

distribution. As Figure 4 shows, if we add the perturbation in the loss distribution, TVaR changes more than VaR, which illustrates that TVaR is more sensitive to perturbation and thus less robust than VaR.

3.4 Score Function-Elicitability

3.4.1 Definition and Properties

Usually, risk measures are calculated based on historical or empirical data. In order to arrive at the best possible point estimate, we have to make appropriate decisions concerning models, methods and parameters. Thus, it is crucial to be able to validate and compare competing estimation procedures.

Let $\mathcal{P}$ be a class of probability measures on $\mathbb{R}$ with the Borel sigma algebra. We consider a functional

$$\nu : \mathcal{P} \to 2^{\mathbb{R}}, \ P \mapsto \nu(P) \subseteq \mathbb{R},$$

where $2^{\mathbb{R}}$ denotes the power set of $\mathbb{R}$.

Formally, a statistical functional is a potentially set-valued mapping from a class of probability distributions. We want to make sure that we can find an optimal point of this function. Here is the definition of the consistency for a scoring function.
Definition 3.6. (Consistency (Gneiting, 2011)) A scoring function $s : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ is consistent for the functional $\nu$ relative to the class $\mathcal{P}$, if

$$E_{P}s(t, Y) \leq E_{P}s(x, Y)$$

for all $P \in \mathcal{P}$, all $t \in \nu(P)$, and all $x \in \mathbb{R}$, where $Y$ is a real-valued random variable and has distribution $P$. It is strictly consistent if it is consistent, and equality implies that $x \in \nu(P)$. Given the scoring function $s$ is consistent for a functional $\nu$, an optimal forecast $\hat{x}$ for $\nu(P)$ is given by

$$\hat{x} = \arg \min_{x} E_{P}s(x, Y).$$

After defining consistency in robust measure, we now turn to the definition of elicitability.

Definition 3.7. (Elicitability) A functional $\nu$ is elicitable relative to the class $\mathcal{P}$, if there exists a scoring function $s$ which is strictly consistent for $\nu$ relative to $\mathcal{P}$.

The most prominent example concerning risk management may be VaR, which is essentially a quantile and as such elicitable. However, not all functionals are elicitable, the most striking example in the present context being TVaR. Here we provide Theorem 3.2 and 3.3 to identify if a score function satisfies elicitability.

Theorem 3.2. (Osband, 1985) An elicitable functional $\nu$ has convex level sets in the following sense: If $P_0 \in \mathcal{P}$ and $P_1 \in \mathcal{P}$, and $P^* = pP_0 + (1 - p)P_1 \in \mathcal{P}$ for some $p \in (0, 1)$, then $t \in \nu(P_0)$ and $t \in \nu(P_1)$ imply $t \in \nu(P^*)$.

Theorem 3.3. (Kou and Peng, 2014) Let $D^*$ be the class of distributions with finite support and $\rho$ be a distortion risk measure with distortion function $h \in H$ defined by

$$\rho(F) = \int_{-\infty}^{0} (h(1 - F(x)) - 1)dx + \int_{0}^{\infty} h(1 - F(x))dx,$$

whose restriction to $D^*$ has convex level sets. Then, $h$ is either the identity on $[0, 1]$ or it is the form

$$h(x) = \begin{cases} 0, & x \in [0, \alpha), \\ c, & x = \alpha, \\ 1, & x \in (\alpha, 1], \end{cases}$$

for some $\alpha, c \in [0, 1]$. If $\alpha = 0$ or $\alpha = 1$, then $c = 0$ or $c = 1$, respectively.
Table 2: Score of each estimator

<table>
<thead>
<tr>
<th></th>
<th>True Estimation Score</th>
<th>Mean Estimation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S(x, y) = (x - y)^2$</td>
<td>4082.363</td>
<td>4142.571</td>
</tr>
<tr>
<td>$S(x, y) =</td>
<td>x - y</td>
<td>$</td>
</tr>
<tr>
<td>$S(x, y) =</td>
<td>(x - y)/x</td>
<td>$</td>
</tr>
</tbody>
</table>

3.4.2 Competing estimate procedures

By referring to the score ranking of scoring function, we can compare different risk estimators. Suppose that in $n$ estimate senarios we have point estimations $x_1^{(k)}, \ldots, x_n^{(k)}, k = 1, \ldots, K$, and realizing observations $y_1, \ldots, y_n$. The index $k$ numbers the $K$ competing forecast procedures. We can rank the procedures by their average scores

$$s^{(k)} = \frac{1}{n} \sum_{i=1}^{n} s(x_i^{(k)}, y_i).$$

The consistency of the scoring rule for the functional $\nu$ ensures that accurate forecasts of $\nu(P)$ are rewarded. On the contrary, evaluating point forecasts with respect to some poorly selected scoring function, which is not consistent for $\nu$, may lead to grossly misguided conclusions about the quality of the estimate. We will provide a small simulation study by Example 3.2 to illustrate it.

Example 3.2. We generate a series where $Y_t = Z_t^2$ and $Z_t$ is a Garch(1,1) series with $\alpha = 0.2$, $\beta = 0.8$ and $\omega = 0.05$. Here we want to estimate the value of $Y_t$ series.

Based on the generating process, the true predict level is $\mathbb{E}(Y_t) = \sigma_t^2$. In this simulation study, we will compare the ‘true estimation’ and ‘mean estimation’ with score functions of squared error $S(x, y) = (x - y)^2$, absolute error $S(x, y) = |x - y|$, and relative error $S(x, y) = |(x - y)/x|$ as in Figure 5.

Based on Figure 5, we can find that the true estimator fits better than the mean estimator. However, if we choose the inappropriate score function, we may misjudge the fitted level of each model and make a wrong decision.

The squared error score function and absolute error score function illustrates our intuition that true estimator is a better estimator to predict the value of series. However, if we choose $S(x, y) = |(x - y)/x|$ as our score function to select model, we may make a wrong decision.
Figure 5: True Estimation and Mean Estimation

to select mean estimator as our measure function, which may lead to grossly misguided conclusions.

3.5 Summary

To end this section, we provide Table 3 to summarize the properties of VaR and TVaR.
Table 3: Some Properties of VaR and TVaR

<table>
<thead>
<tr>
<th>Properties</th>
<th>$VaR_\alpha$</th>
<th>$TVaR_\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coherence</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Convexity</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Comonotonic additivity</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Law Invariance</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Qualitative Robustness</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Robustness with the Wasserstein distance</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bounded Pertubation Sensitivity</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Elicitability</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
4 Robustness in Other Fields

4.1 Cyber Robustness

Cyber Robustness metrics quantify the possibility of success and time taken for an ICT system resisting to attackers who arrange the sequence of attacks for their own privileges and goals. The fundamental evaluation has adopted the Haruspex suite result; it is forecasting a system through simulation of the interaction between attackers and the system which resulted to robuster design and metrics computation. The model builds the statistical samples to access the ICT risk by using the Monte Carlo method in the computation of probability given their measurable parameters in the model, such as the vulnerabilities in each system module, the attacks, and their success probability. Besides, the attacker also has been modeled as at and included two important properties: the selection strategy of at and the look ahead, sg, a non-negative integer presents the selection way of attacks by the attacker in its escalations.

Baiardi et al. (2016) suggested three types of metrics for this evaluation. First is by using the security stress that calculates the probability of an attacker to reach his goal given a time interval and number of attacks in an escalation. The probability distribution of stress is defined by

\[ Str_{at,sg}^S(t) = PrSucc_{at,sg}^S(t), \]

where \( PrSucc_{at,sg}^S(t) \) is the sum for all the possible escalations of the probability each escalation is successful at \( t \). It is also identified to increases with \( t \) as larger value of \( t \) allows a larger number of attack failures so the inverse of stress. Note that

\[ Sur_{at,sg}^S(t) = 1 - Str_{at,sg}^S(t) \]

presents the probability that \( S \) survives to the attacks of \( at \) to reach a goal in \( sg \). We can also compute the stress of a set of shared mission attacker by using the model:

\[ Str_{at,sg}^S(t) = \max\{ at \in sa, Str_{at,sg}^S(t) \}, \]

which is the largest stress of its attackers.

Here, we have two kinds of distributions: the \( AvLoss \) measures the expected loss at the certain time given discrete attackers corresponding rights using the first-order derivative of stress as the probability. The average loss of for an agent with a goal is defined by the model

\[ AvLoss_{sg,sg}^S(t) = \int_{t' \in [0, ..., t]} Str_{sg,sg}^S(t')Imp_{sg,sg}^S(t - t')dt', \]

28
while in general scenarios, the sum of the average losses due to each agent is

$$AvLoss^S_{sag, sg}(t) = \sum_{agi \in sag} AvLoss^S_{agi, sg}(t)$$

where $sag = \{a_1, \ldots, a_f\}$ and each agent in $sag$ has the goals in $sg = \{g_1, \ldots, g_n\}$.

The next one is $CyVar$ that assesses loss while applying the VaR model. As VaR widely used for security investment, this metric evaluation chooses to focus on the same perspective; it estimates the chances of losing money given the time, the confidence level, and the loss amount. $CyVar$ consists of two models: for one agent with alternative goals, approximation of largest probability loss such in the given model

$$CyVar^S_{ag, sg}(\nu, t) = \max\{CyVar^S_{ag, gi}(\nu, t), gi \in sg\}$$

and the model for numbers of agents included. Conclusively, $CyVar$ gives more accurate evaluation compared than $AvLoss$.

4.2 Flood and Drought Risk Management

Climate change has increased the uncertainty of the likeliness of floods and droughts occurring. Robustness helps to give more accurate flood risk analysis through assessing a large range of possible floods in the analysis. Robustness in the drought risk analysis adds to a wider definition of drought magnitude, giving us more supply reliability. It also allows us to take into account worst-case climate change scenarios.

In conjunction to that, a system robustness analysis has been developed, by exploring the characteristics of the event and measuring the severity and impact, to access the ability of the system. It is important to characterize the system with resilience, the system’s ability to recover from response to a disturbance, and resistance, the endurance ability without responding, which both give the analysis on sensitivity of the system in detail. Researchers have drawn the relationship between the event severity and corresponding response. In general, the robustness model includes few important criteria in the analysis such as the resistance threshold, the proportionality of the relationship, and the manageability of the system.

This developed robustness model has been successfully applied in the flood risk system.
In the system analysis, while the protection standard is chosen as the resistance threshold, it is found that the sudden change in river flow does not significantly contribute to the flood’s impacts. Thus, this result has enable us to manage the impact whilst below the critical level even for wider range of magnitudes.

Mens and Kirin (2015) demonstrated that the robustness criteria have additional value compared to the more traditional decision making criteria, based on single-value risk, by performing a case study on the IJssel River valley in the Netherlands. They access the flood risk by calculating the water level probability distribution per breach location. The value is then integrated into a fragility curve giving us the flood probability distribution of

\[ P_k = \int f_k(h) \cdot PC_k(h) \cdot dh \]

which \( P_k \) is the flood probability of location \( k \), \( f_k(h) \) is the water level probability density function and \( PC_k(h) \) is the conditional probability of embankment breaching:

\[ PC_k(h) = \Phi(\mu = m; \sigma = 0.2). \]

Using the maximum flood depth maps as input for the damage model, the flood risk calculation of the IJssel valley combines flood probabilities, which is defined by a normal distribution where \( Pr(Z < 0) \) means failure, and consequences of at each of location; the risk is known by determining the area under the normal curve:

\[ \text{Flood Risk} = \mathbb{E}[D] = \int P(D) \cdot D \cdot dD = \int F(D) \cdot dD, \]

where \( D \) is flood damage [EUR], \( F(D) \) is the probability density of the damage, \( P(D) \) is the probability of one damage scenario and \( \mathbb{E}(D) \) is the expected value of the damage. In addition, the robustness analysis was completed by applying the Monte Carlo approach; it resulted in proving that the resistance threshold is similar in all configurations, because it does not significantly depend on assumptions about discharge variability and climate change.

Meanwhile, Mens et al. (2015) demonstrated the application of the robustness system in the drought risk analysis on a case inspired by the Oologah reservoir in Oklahoma, United States. In addition, they used different criteria in the analysis like Demand reduction, Hedging, and Reservoir expansion. First, the drought volume is determined by the equation:

\[ \text{Volume} = \sum_{t_{on}}^{t_{off}} (R_t - Q_t) = K_{t_{off}}. \]
Then, the impact of the event was determined in terms of change in welfare through the willingness to pay; it was estimated by a total loss function in US Dollar that includes the amount of available water, the baseline water use, water rate and price elasticity. Finally, the robustness criteria was scored using the response curve. The demand reduction is preferable over the supply increase on the supply reliability in the side of robustness. Additionally, focusing on demand reduction allows us to deal with similar extreme conditions. Nevertheless, it is recommended to put the framework into testing different systems or types of floods (Mens et al., 2015).

4.3 The Robustness of Power Grids

The significant dependency on electric power grids for electric supply in a country leads to higher control of risk or failure limitation. According to the North American Electrical Reliability Council (NERC), data shows that large blackouts happen more frequently than expected due to line overloads or failure of any single transmission line. It is important to analyze and minimize possible risk, and this could be done by a robustness metric of a power transmission grid with respect to cascading failures (Koç et al., 2013).

A new notion of network robustness, there is only one metric to evaluate it that was induced by random failures. Koç et al. (2013) confirmed that there are two important factors in the robustness evaluation. The structure and the operative structure of a network. In addition, the metric also depends on two main concepts: the electrical nodal robustness, and the electrical node significance which uses an entropy-based approach and a nodal centrality measure.

Koç et al. (2013) developed a model for each factor of the cascading effects. The line overloads in power networks is modeled using the complex networks approach. The system includes the power grid as a graph, the line flows across the grid analysis, and estimation of the cascading damages which is defined with the linear equation

\[ P_i = \sum_{j=1}^{d} f_{ij} = \sum_{j=1}^{d} b_{ij} \theta_{ij} \]

where \( P_i \) is the real power flow at node \( i \), and \( d \) is the degree of node \( i \). In specific, the capacity of a line was introduced as the maximum transportable power flow by the line, \( C_i \) which is proportional to its initial load, \( L_i(0) \), given the tolerance parameter of line \( i \), \( \alpha_i \), i.e. \( C_i = \alpha_i L_i(0) \).
Meanwhile, Koç et al. (2013) has quantified the cascading failures by the metrics:

- **Demand Survivability (DS):** The fraction of the satisfied power demand after a cascading failure occurs in a network.

- **Link Survivability (LS):** The fraction of lines that are still in operation after a cascading failure with definition

\[
LS = \frac{L'}{L}
\]

where \(L\) is the total number of links and \(L'\) the number of links operational after a cascading failure.

- **Capacity Survivability (CS):** The fraction of capacity of the operational lines after a cascading failure with definition

\[
CS = \frac{\sum_{i=1}^{L'} C_i}{\sum_{j=1}^{L} C_j}
\]

with \(C\) being the sum of the capacity of the links in the network and \(C'\) the sum of the capacity after a cascading failure.

However, operators need considerable computational power and time for large networks. The paper introduced a robustness metric, \(R_{CF}\), to encounter the issue. This metric relies on two main concepts: electrical nodal robustness and electrical node significance. First, the electrical nodal robustness is the aggregate value that represents the ability to withstand cascades of link overload failures as well as takes flow dynamics and topology effects on network robustness into account with equation

\[
R_{n,i} = -\sum_{i=1}^{L} \alpha_i p_i \log p_i,
\]

where \(p_i\) stands for values in the distribution under consideration. On the other side, electrical node significance suggests the impact of a particular node as

\[
\delta_i = \frac{P_i}{\sum_{j=1}^{N} P_j},
\]

with \(P_i\) standing for total power distributed by node \(i\), and \(N\) refers to the number of nodes in the network. Finally, the computation of network robustness metric,
was done by adding all individual contributions of each node in the network with respect to cascading failures.

This shows how the computation does not include expensive tasks such as simulation of cascading failures. The effectiveness of this robustness metric has been verified via experimental developed models and shown applicable on different cases including IEEE test systems and UCTE networks. It is shown that the computation is parallel and how the computation does not include expensive tasks such simulation of cascading failures. The properties of $R_{CF}$ have allowed us to use it as a real-time measure while monitoring and optimizing it dynamically.

4.4 Robustness Evaluation for Bio-manufacturing

Both drug development and manufacturing are dependent upon economic and regulatory factors that can impact industry decision-making towards cost-effective and value-potential alternatives for areas such as: bioprocess and facility design, capacity sourcing and portfolio selection. Farid (2013) summarized systematic approaches and evaluations established at University College London (UCL) in addressing this issue; including some techniques such as: process economics, simulation, risk analysis, optimization, operations research and multivariate analysis. This evaluation develops a model to estimate cost of goods and other cost metrics. The model also consolidates bioprocess economics, manufacturing logistics through discrete-event simulations, and uncertainties through Monte Carlo simulation to assess the robustness and assist in the decision-making process.

One of the applications of this model is a case study for a company which was required to make a decision on a pipeline of monoclonal antibodies. The decision was either to invest in the disposable facility, the traditional stainless-steel based one or go for a hybrid option at the 200-L scale (Farid, 2013). It used the technique of probabilistic additive weighting to consolidate the trade-offs and uncertainties in the input. Then, it would standardize the financial and operational score into a common dimensionless scale that also indicates the intrinsic risk.

The results show that the preference is for the hybrid option, followed by the disposable and the stainless steel option for earlier stage material. While for later stage materials, the
relative rankings may give a different result. This is due to the former’s operational score relying on other scores, which are very important to that stage, such as the construction time, project throughput, and operational flexibility. In the conclusion, Farid (2013) suggested that the given support tool has crystallized the trade-offs and uncertainties involved, as well as coming up with clear financial, operational and risk metric evaluations. This has allowed the tool being applied across different company departments during both building and analysis.

4.5 Robust Optimization Formulations in Water Resource Systems Planning

Robust optimization (RO) formulations in the water resources planning is important standardize the uncertainty analysis as well as allowing for better evaluation and regulation of various risks of poor system performance. These risks are solution robustness, reliability, vulnerability and sustainability. Previously, RO has been modeled by a single metric showing inadequate importance compared to RO evaluation through post-processing given a wider selection of performance metrics. The failure of the former model is the lacking of the most basic shortcoming, which is the operational trade-off. This is proven via analysis of the trade-offs between solution robustness and its feasibility over all possible scenarios (Ray et al., 2014).

Ray et al. (2014) has investigated the robustness of different models that have been developed over time. One of the earliest formulations of this problem was completed by Lund and Israel in 1995. They computed the minimization of the expected total of direct and indirect costs. In this first model, the potential water availability and usage was used as the random input parameter with related probabilities. The model is defined in the following two-stage stochastic nonlinear program:

$$
\min Z_1 = c_c Q + \sum_{s \in S} \sum_{r \in R} p_r p_s [c_0 U_{q_{rs}} + c_t U_{t_{rs}} + \eta(U_{s_{rs}})^\gamma],
$$

where $Z_1$ is the objective function value, total cost (direct and indirect). $c_c$ is the unit capital cost of desalination plant. $Q$ is the capacity of desalination plant. $p_r$ is the probability of water requirement event $r$. $p_s$ is the probability of supply event $s$. $c_0$ is the amortized unit operation and maintenance cost of desalination plant. $U_{q_{rs}}$ is the capacity of desalination plant actually used. $c_t$ is the unit cost of water transfer. $U_{t_{rs}}$ is the quantity of transfer water purchased. $U_{s_{rs}}$ is the quantity of water shortage. $\eta(U_{s_{rs}})^\gamma$ is the nonlinear cost of
water shortage.

Although the previous model gives the least-cost solution to the problem, it still requires a multi-objective approach when it is high dimensional. The next model shows the multi-objective two-stage robust optimization model (MO-RO) by Watkins and McKinney. It is defined by a model of standard deviation of possible water-related cost in the future which has been identified as not monotonically increasing. This model has been developed to improve the rationality of second-stage decision process making. It also offers to simultaneously control the sensitivity of solutions through the extension of stochastic programming to a multi-objective optimization framework which manages to reflect risk-averse behavior in the objective function. Therefore, the relationship between solution robustness and its feasibility is fairly important for optimization.

Nevertheless, the second model is more vulnerable to shortage range in a specified range. There is a newer model that consists of three alternative MO-RO formulation, that fundamentally penalizes the square of positive deviations from a fixed target cost. This model gives a better result in terms of smaller standard deviation and direct cost compared to Model 1. Besides, it can reduce the dependability on water usage during drier years by using more excess capacity and accepting a fair amount of the expected cost.
5 Conclusion

In this paper, we studied the various risk measures in the literature and the robustness metrics for VaR and TVaR. It is worth to investigate the properties of robustness metrics for different risk measures and apply the methods in other fields to the framework.
References


