# Robustness and sensitivity analysis of risk measurement procedures*

Rama CONT[1], Romain DEGUEST [1,2] and Giacomo SCANDOLO[3]

1) IEOR Dept, Columbia University, New York.
2) CMAP, Ecole Polytechnique, France.
3) Dipartimento di Matematica per le Decisioni, Università di Firenze, Italia.

Financial Engineering Report No. 2007-06
Center for Financial Engineering, Columbia University.

**Abstract**

Measuring the risk of a financial portfolio involves two steps: estimating the loss distribution of the portfolio from available observations and computing a "risk measure" which summarizes the risk of the portfolio. We define the notion of "risk measurement procedure", which includes both of these steps and study the robustness of risk measurement procedures and their sensitivity to a change in the data set. After introducing a rigorous definition of 'robustness' of a risk measurement procedure, we illustrate the presence of a conflict between subadditivity and robustness of risk measurement procedures. We propose a measure of sensitivity for risk measurement procedures and compute the sensitivity function of various examples of risk estimators used in financial risk management, showing that the same risk measure may exhibit quite different sensitivities depending on the estimation procedure used. Our results illustrate in particular that using historical Value at Risk leads to a more robust procedure for risk measurement than recently proposed alternatives like CVaR. We also propose other risk measurement procedures which possess the robustness property.

*We thank Hans Föllmer, Peter Bank, Paul Embrechts, Gerhard Stahl and seminar participants at QMF 2006 (Sydney), Humboldt University (Berlin), Torino, Lecce, INFORMS Applied Probability Days (Eindhoven), Cornell ORIE seminar and the Harvard Statistics Seminar for helpful comments. This project has benefited from partial funding by the European Research Network "Advanced Mathematical Methods for Finance" (AMAMEF).

Keywords: robust statistics, coherent risk measures, law invariant risk measures, Value at Risk, expected shortfall.

## Contents

## 1 Introduction

One of the main purposes of mathematical modeling in finance is to quantify the risk of financial portfolios. In connection with the widespread use of Value at Risk and related risk measurement methodologies and the Basel committee guidelines for risk-based requirements for regulatory capital, methodologies for measuring of the risk of financial portfolios have been the focus of recent attention and have generated a considerable theoretical literature [1, 2, 3, 9, 8, 10]. In this theoretical approach to risk measurement, a risk measure is represented as a map assigning a number (a measure of risk) to each random payoff. The focus of this literature has been on the properties of such maps and requirements for the risk measurement procedure to be coherent, in a static or dynamic setting.

An implicit starting point of these developments is that the underlying probability measure describing market events is *known*. This is crucial since most risk measures such as Value-at-Risk or Expected Shortfall are defined as functionals of the distribution of the considered payoff. In applications, however, this probability distribution is unknown and should be estimated from (historical) data as part of the risk measurement procedure. Thus, in practice, measuring the risk of a financial portfolio involves two steps: estimating the loss distribution of the portfolio from available observations and computing a risk measure of this distribution which summarizes the risk of the portfolio. While these two steps have been considered and studied separately, they are intertwined in applications and an important criterion in the choice of a risk measure is the availability of good estimation procedures. In order to study the interplay of a risk measure and its estimation method used for computing it, we define the notion of *risk measurement procedure*, as a two–step procedure which maps a couple $(D_n, X)$, where $X$ is a payoff and $D_n$ a data set on which the estimation is based, to a risk estimate $\widehat{\rho}(X)$ for $X$. This quantity is supposed to estimate the "abstract" risk measure $\rho(X)$ estimated from the data set $D_n$. This distinction between an "abstract" risk measure and its estimation procedure allows us to study the interplay between the specification of risk measures and the choice of the estimator.

Once we have acknowledged the uncertainty in the risk estimates resulting either from estimation or mis-specification errors in the loss distribution, it is natural to examine the sensitivity of the results with respect to these errors. Of particular concern is the **robustness** of the risk measurement: the estimator for the portfolio's risk is said to be robust if small variations in the loss distribution –resulting either from estimation or mis-specification errors– should result in small variations in the estimator.

We propose a rigorous approach for examining these issues, using tools from robust statistics. We introduce a qualitative notion of 'robustness' for a risk measurement procedure and a way of quantifying it via sensitivity functions. Using these tools we show that there is a conflict between coherence (more precisely, the sub-additivity) of a risk measure and the robustness, in the statistical sense, of its commonly used estimators. This consideration goes against the traditional arguments for the use of coherent risk measures in risk measurement and deserves discussion. We complement this abstract result by computing measures of sensitivity, which allow to quantify the robustness of various risk measures with respect to the data set used to compute them. In particular, we show that the same "risk measure" may exhibit quite different sensitivities depending on the estimation procedure used. These properties are studied in detail for some well known examples of risk measures: Value at Risk, Expected Shortfall/ CVaR [1, 16, 17] and the class of spectral risk measures introduced by Acerbi [2]. Our results illustrate in particular that using historical Value at Risk instead of alternative risk measures, suggested in the recent theoretical literature, leads to a more robust procedure for risk measurement.

The article is structured as follows. Section 2 recalls some basic notions on distribution-based risk measures. In Section 3 we establish the distinction between an abstract risk measure and a *risk measurement procedure*. We then show that a risk measurement procedure applied to a data set can be viewed as the application of an *effective* risk measure to the empirical distribution obtained from this data. We give examples of effective risk measures associated to various risk measurement procedures.

Section 4 defines the notion of *robustness* for a risk measurement procedure and examines whether this property holds for commonly used risk measurement procedures. We show in particular that there exists a conflict between the sub-additivity of a risk measure and robustness of its estimation procedure.

In section 5 we define the notion of *sensitivity function* for a risk measure and we compute sensitivity functions for some commonly used risk measurement procedures. The behavior of these sensitivity functions allow us to compare different risk measurement procedures in terms of their sensitivity to a change in the underlying data set.

We discuss in section 7 some implications of our findings for the choice of risk measures and the design of risk measurement procedures in finance.

## 2 Risk measures

### 2.1 Notations

We introduce here some basic notations to be used throughout the paper. We shall denote $\mathcal{D} = \mathcal{D}(\mathbb{R})$ the (convex) set of cumulative distribution functions (cdf) on $\mathbb{R}$. The weak topology on $\mathcal{D}$ is defined by the Levy distance:

$$d_L(F,G) \triangleq \inf\{\varepsilon > 0 \,:\, F(x-\varepsilon) - \varepsilon \leq G(x) \leq F(x+\varepsilon) + \varepsilon \;\forall x \in \mathbb{R}\},$$

or by the equivalent Prohorov distance [12]. The upper and lower quantiles of $F \in \mathcal{D}$ of order $\alpha \in (0,1)$ are defined, respectively, by:

$$q_\alpha^+(F) \triangleq \inf\{x \in \mathbb{R} \,:\, F(x) > \alpha\} \;\text{ and }\; q_\alpha^-(F) \triangleq \inf\{x \in \mathbb{R} \,:\, F(x) \geq \alpha\}.$$

Plainly $q_\alpha^+(F) \geq q_\alpha^-(F)$, but, for any fixed $F$, equality holds for almost all $\alpha \in (0,1)$. Moreover, the maps $\alpha \mapsto q_\alpha^+(F), q_\alpha^-(F)$ are increasing, thus measurable; therefore, integrals involving these maps are well defined.
For $p \geq 1$ let $\mathcal{D}^p$ be the set of distributions having finite $p$-th moment and by $\mathcal{D}^p_-$ the set of distributions whose left tail has finite $p$-th moment. We denote $\mu(F)$ the mean of $F \in \mathcal{D}^1$ and $\sigma^2(F)$ the variance of $F \in \mathcal{D}^2$. For any $n \geq 1$ and any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, let

$$F^{\mathbf{x}}(x) \triangleq \frac{1}{n}\sum_{i=1}^n I_{\{x \geq x_i\}}$$

the associated empirical cdf; $\mathcal{D}_{\text{emp}}$ will denote the set of all empirical cdf.
If $X$ is a random variable defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then $F_X \in \mathcal{D}$ denotes the distribution of $X$ under $\mathbb{P}$, i.e. $F_X(x) = \mathbb{P}(X \leq x)$.

### 2.2 Coherent risk measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space representing market scenarios and let $L^0$ denote the space of random variables defined on it. Each random variable $X \in L^0$ represents the "Profit and Loss" (P&L) deriving from the holding of a *portfolio* over a specified horizon; in our framework, negative values for $X$ correspond to losses, as in [3] and subsequent works, but unlike most literature in actuarial risks. Consider a convex cone $V \subset L^0$ containing all the constant r.v. Generally speaking, a *risk measure* on $V$ is a map $\rho : V \to \mathbb{R}$ assigning to each P&L $X \in V$ its degree of riskiness. Artzner et al [3] formulated a set of requirements, known as *coherency axioms* that monetary measures of risk should verify:

**Definition 1** (Coherent risk measure)**.** A risk measure $\rho : V \to \mathbb{R}$ is *coherent* if it is $(X, Y \in V)$:

1. *Monotone (decreasing).* $\rho(X) \le \rho(Y)$ provided $X \ge Y$
2. *Cash* (or *translation*) *invariant.* $\rho(X + c) = \rho(X) - c$ for any $c \in \mathbb{R}$.
3. *Subadditive.* $\rho(X + Y) \le \rho(X) + \rho(Y)$
4. *Positive homogeneous.* $\rho(\lambda X) = \lambda \rho(X)$ for any $\lambda \in \mathbb{R}_+$:

### 2.3 Distribution-based risk measures

Most risk measures used in practice are distribution-based, i.e. they depend only on the distribution of the portfolio gain/loss.[1]

**Definition 2** (Distribution-based risk measure)**.** A risk measure $\rho : V \to \mathbb{R}$ is said to be *distribution-based* if $\rho(X) = \rho(Y)$ whenever $F_X = F_Y$.

Putting $\rho(F_X) \triangleq \rho(X)$, we can represent a distribution-based risk measure $\rho$ on $V$ as a map (also denoted $\rho$) defined on the set of loss distributions $\mathcal{D}_\rho = \{F_X \in \mathcal{D} : X \in V\}$.

A relevant class of distribution-based risk measures which includes all examples used in applications is given by

$$\rho_m(F) = -\int_0^1 q_u^+(F)\, m(du), \tag{1}$$

where $m$ is a probability measure on $[0, 1]$. We will denote $\mathcal{D}_m$ the set of distributions for which the integral above is finite. Notice that if the support of $m$ does not contain 0 nor 1, then $\mathcal{D}_m = \mathcal{D}$. For any choice of the weight $m$, $\rho_m$ is monotone, translation invariant and positive homogeneous. The subadditivity of such risk measures can be easily characterized [2, 9, 15]:

**Proposition 1.** *A risk measure $\rho_m$ as in* (1) *is sub-additive (hence coherent) on $\mathcal{D}_m$ if and only if $m$ has a decreasing density: $m(du) = \phi(u)du$ where $\phi$ is a positive decreasing function.*

Three particular cases deserve attention:

[1] In the literature such risk measures are often called "law-invariant", an awkward name which conveys in fact the opposite of what it should.

**Value at Risk** It corresponds to the choice $m = \delta_\alpha$ for a fixed $\alpha \in (0,1)$ (usually $\alpha \leq 5\%$) and therefore is defined as:

$$\mathrm{VaR}_\alpha(F) \triangleq -q_\alpha^+(F). \tag{2}$$

$\mathrm{VaR}_\alpha$ is not subadditive, and its domain of definition is all $\mathcal{D}$.

**Expected shortfall (CVaR)** corresponds to choosing $m$ as the uniform distribution over $(0,\alpha)$, where $\alpha \in (0,1)$ is fixed (again, usually $\alpha =$ 1% or 5%):

$$\mathrm{CVaR}_\alpha(F) \triangleq \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_u(F)\,du. \tag{3}$$

In this case, $\mathcal{D}_m = \mathcal{D}_-^1$, the set of distributions having integrable left tail. Contrarily to VaR, CVaR is a coherent risk measure [1, 2, 9].

**Spectral risk measures** [15, 1, 2] This class of risk measures generalizes CVaR and corresponds to choosing $m(du) = \phi(u)du$, where $\phi$ is a decreasing probability density on $(0,1)$. Therefore:

$$\rho_\phi(F) \triangleq \int_0^1 \mathrm{VaR}_u(F)\,\phi(u)\,du. \tag{4}$$

In view of Proposition 1, spectral risk measures are exactly the risk measures in (1) which are coherent. If $\phi \in L^q(0,1)$ (but not in $L^{q+\varepsilon}$) and $\phi \equiv 0$ around 1, then $\mathcal{D}_m = \mathcal{D}_-^p$, where

$$\frac{1}{p} + \frac{1}{q} = 1$$

# 3 Risk measurement procedures

## 3.1 Estimation of risk measures

Once a risk measure $\rho$ has been chosen, in practice one has to estimate the loss distribution of the portfolio from available data, then compute the risk measure using this distribution. Assume that the portfolio loss can be written as $X = h(\mathbf{Z})$, where $\mathbf{Z} \in \mathbb{R}^K$ is a random vector describing risk factors. The estimation of a risk measure for the portfolio then involves the following steps:

1. Estimating the distribution of the risk factors from observations $\mathbf{z}_1, \ldots, \mathbf{z}_N$ of the vector $\mathbf{Z}$. This estimation may be using nonparametric or parametric methods, leading to an estimated distribution $\widehat{F}_{\mathbf{Z}}$.

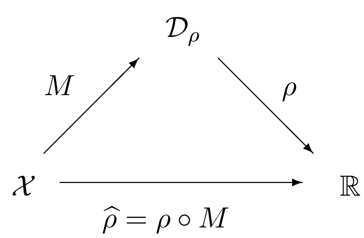
2. Computing the loss distribution of the portfolio from the distribution $\widehat{F}_{\mathbf{Z}}$ of the risk factors. In simple cases (e.g. linear portfolios) it is possible to compute this distribution analytically but in most cases (e.g. portfolios with options) a Monte Carlo simulation may be needed at this stage. The result of this second step is an estimated loss distribution $\widehat{F}_X$.

3. Applying the risk measure to the estimated loss distribution $\widehat{F}_X$, in order to obtain an estimate $\rho(\widehat{F}_X)$.

For ease of presentation, we shall consider the setting where the data are assumed to come from an IID sample of portfolio losses (time series or Monte Carlo simulation), that is $X = Z$. In this setting the procedure consists in:

1. Estimating the loss portfolio distribution from a data set. This estimation step can be represented as a map $M : \mathcal{X} \to \mathcal{D}_\rho$, where $\mathcal{X} = \cup_{n\geq 1}\mathbb{R}^n$ represents data sets.

2. Computing the risk associated to the loss portfolio through a *risk measure* $\rho : \mathcal{D}_\rho \to \mathbb{R}$.




This naturally leads to the notion of *risk measurement procedure*, which integrates these two steps:

**Definition 3** (Risk measurement procedure)**.** A *risk measurement procedure* (RMP) is a couple $(M, \rho)$, where $\rho : \mathcal{D}_\rho \to \mathbb{R}$ is a risk measure and $M : \mathcal{X} \to \mathcal{D}_\rho$ an estimator for the loss distribution.

From the operational viewpoint, what really matters is the composition $\widehat{\rho} = \rho \circ M : \mathcal{X} \to \mathbb{R}$ which gives us a recipe to directly compute an estimate for the risk measure out of the data-set (see diagram). We will call the map $\widehat{\rho}$ the *risk estimator* for $\rho$ associated to the estimation method $M$.

Given a risk measure $\rho$, various estimation procedures can lead to different risk measurement procedures. We now describe some important examples.

### 3.2 Historical risk estimators

Given a data set $\mathbf{x} \in \mathcal{X} = \cup_{n \geq 1} \mathbb{R}^n$, the historical risk estimator $\widehat{\rho}^h$ associated to a risk measure $\rho$ is the estimator obtained by applying the risk measure $\rho$ to the empirical loss distribution (sample cdf) $F^{\mathbf{x}}$:

$$\widehat{\rho}^h(\mathbf{x}) = \rho(F^{\mathbf{x}}).$$

For a risk measure $\rho_m$, defined in (1) we can easily compute:

$$\widehat{\rho}^h_m(\mathbf{x}) = \rho_m(F^{\mathbf{x}}) = -\sum_{i=1}^{n} w_{n,i}\, x_{(i)},$$

where $x_{(k)}$ is the $k$-th least element of the set $\{x_i\}_{i \leq n}$, $w_{n,i} \triangleq m\left(\frac{i-1}{n}, \frac{i}{n}\right]$ for $i = 1, \ldots, n-1$, and $w_{n,n} = m\left(\frac{n-1}{n}, 1\right)$. It follows that:

**Example 3.1.** The historical estimator of $\mathrm{VaR}_\alpha$, $\alpha \in (0,1)$ is:

$$\widehat{\mathrm{VaR}}^h_\alpha(\mathbf{x}) = -x_{(\lfloor n\alpha \rfloor + 1)}, \tag{5}$$

where $\lfloor a \rfloor$ denotes the integer part of $a \in \mathbb{R}$.

**Example 3.2.** The historical estimator of $\mathrm{CVaR}_\alpha$, $\alpha \in (0,1)$ is:

$$\widehat{\mathrm{CVaR}}^h_\alpha(\mathbf{x}) = -\frac{1}{n\alpha}\left(\sum_{i=1}^{\lfloor n\alpha \rfloor} x_{(i)} + x_{(\lfloor n\alpha \rfloor + 1)}(n\alpha - \lfloor n\alpha \rfloor)\right) \tag{6}$$

**Example 3.3.** The historical estimator of the spectral risk measure $\rho_\phi$ associated to $\phi : [0,1] \to [0,+\infty)$ is given by:

$$\widehat{\rho}^h_\phi(\mathbf{x}) = -\sum_{i=1}^{n} w_{n,i}\, x_{(i)}, \tag{7}$$

where, for any $n \geq 1$ and $i = 1, \ldots, n$

$$w_{n,i} = \int_{(i-1)/n}^{i/n} \phi(u)\, du.$$

Historical risk estimators thus belong to the class of L-estimators in the sense of Huber [12].

### 3.3 Parametric risk estimators

In the parametric approach, we choose a parametric family $\mathcal{D}_\Theta = (F_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta} \subset \mathcal{D}$ of loss distributions, where $\Theta \subset \mathbb{R}^K$ ($K \geq 1$) is the set of (real) parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ and we assume that the loss distribution lies in this family. For a risk measure $\rho : \mathcal{D}_\rho \to \mathbb{R}$ (with $\mathcal{D}_\Theta \subseteq \mathcal{D}_\rho$), we can map the parameter of the model to the risk estimate by

$$r(\boldsymbol{\theta}) = \rho(F_{\boldsymbol{\theta}}), \quad \boldsymbol{\theta} \in \Theta$$

Then, an estimator of $\boldsymbol{\theta}$ is chosen, i.e. a map $\widehat{\boldsymbol{\theta}} : \mathcal{X} \to \Theta$ assigning to each data set $\mathbf{x} \in \mathcal{X}$ a value $\widehat{\boldsymbol{\theta}}(\mathbf{x}) \in \Theta$ of the parameter(s). The risk estimator associated to the parametric family $(F_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Theta}$ is then defined as

$$\widehat{\rho}(\mathbf{x}) = \rho(F_{\widehat{\boldsymbol{\theta}}(\mathbf{x})}) = r(\widehat{\boldsymbol{\theta}}(\mathbf{x})), \quad \mathbf{x} \in \mathcal{X}.$$

The main example we will study is the maximum likelihood estimator. More generally one can consider the class of M-estimators, obtained by solving a minimization problem of the form:

$$\widehat{\boldsymbol{\theta}}(\mathbf{x}) = \text{arg.max}_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} \Psi(x_i, \boldsymbol{\theta}), \qquad \mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}, \tag{8}$$

for some function $\Psi = \Psi(x, \boldsymbol{\theta})$.
If $F_{\boldsymbol{\theta}}$ admits a strictly positive density $f_{\boldsymbol{\theta}}$ and we take $\Psi(x, \boldsymbol{\theta}) = \log f_{\boldsymbol{\theta}}(x)$ in (8), then we obtain the Maximum Likelihood Estimator (MLE) of $\boldsymbol{\theta}$, denoted by $\widehat{\boldsymbol{\theta}}_{MLE}$.

We will focus in what follows on Maximum Likelihood Estimators for location-scale families. This includes, for instance, the variance-covariance approach to estimate Value at Risk.

Let $F_\star \in \mathcal{D}$ be a fixed distribution; the *location-scale* family associated with $F_\star$ is defined as:

$$\mathcal{D}_{F_\star} \triangleq \{F_{\mu,s} \,:\, \mu \in \mathbb{R},\, s > 0\}, \quad \text{where } F_{\mu,s}(x) \triangleq F_\star\left(\frac{x-\mu}{s}\right).$$

Observe that if $F_\star$ is the distribution of some $X$, then $F_{\mu,s}$ is the distribution of $s\,X + \mu$. It easily turns out that $F_{\mu,s}$ has finite mean (variance) if and only if $F_\star$ does. If $F_\star \in \mathcal{D}_2$, we choose to be centered with unit scale parameter: $F_\star = F_{0,1}$. $F_{\mu,s}$ has a density if and only if $F_\star$ has a density and in this case we have:

$$f_{\mu,s}(x) = \frac{1}{s} f_\star\left(\frac{x-\mu}{s}\right).$$

A relevant subfamily is obtained by fixing $\mu = \mu_0$ (in practice, often $\mu_0 = 0$) thus obtaining the *scale* family $\mathcal{D}_{F_\star,\mu_0} \triangleq \{F_{\mu_0,s} \,:\, s > 0\}$. Two important location-scale (or scale) families of (symmetric) distributions that we will study are:

- the Gaussian family where

$$f_\star(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

- the Laplace or double exponential family where

$$f_\star(x) = \frac{1}{2} \exp\left(-|x|\right).$$

For a location-scale family derived from a density $f_\star$, the MLE $\widehat{\mu}_{MLE}$ and $\widehat{s}_{MLE}$ of $\mu$ and $s$ are defined by the equations:

$$\begin{cases} \sum_{i=1}^n \frac{f'_\star\left(\frac{x_i-\widehat{\mu}_{MLE}}{\widehat{s}_{MLE}}\right)}{f_\star\left(\frac{x_i-\widehat{\mu}_{MLE}}{\widehat{s}_{MLE}}\right)} = 0 \\ \sum_{i=1}^n \left[\left(\frac{x_i-\widehat{\mu}_{MLE}}{\widehat{s}_{MLE}}\right) \frac{f'_\star\left(\frac{x_i-\widehat{\mu}_{MLE}}{\widehat{s}_{MLE}}\right)}{f_\star\left(\frac{x_i-\widehat{\mu}_{MLE}}{\widehat{s}_{MLE}}\right)} + 1\right] = 0 \end{cases}$$

Let $\rho$ be a translation invariant and homogeneous risk measure, for instance any risk measure $\rho_m$ of type (1) such as $\mathrm{VaR}$, $\mathrm{CVaR}$ or any spectral risk measure. For the location-scale family $\mathcal{D}_{F_\star}$, we have

$$\rho(F_{\mu,s}) = -\mu + c\, s, \quad \forall F_{\mu,s} \in \mathcal{D}_{F_\star},$$

where $c = \rho(F_\star)$ does not depend on $\mu$ and $s$. As a consequence the ML risk estimator of $\rho$ under the family $\mathcal{D}_{F_\star}$ is

$$\widehat{\rho}_{MLE}(\mathbf{x}) = -\widehat{\mu}_{MLE}(\mathbf{x}) + c\, \widehat{s}_{MLE}(\mathbf{x}).$$

**Example 3.4** (ML risk estimators for a Gaussian family)**.**
Consider the Gaussian scale-location family, which depends on the parameters $\boldsymbol{\theta} = (\mu, \sigma) \in \Theta = \mathbb{R} \times \mathbb{R}_+$. The ML estimators $\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma})$ are the "natural" ones:

$$\widehat{\mu}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n x_i, \quad \widehat{s}(\mathbf{x}) = \widehat{\sigma}(\mathbf{x}) = \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - \widehat{\mu}(\mathbf{x}))^2}. \tag{9}$$

The resulting usual Gaussian ML risk estimators are:

$$\widehat{\rho}(\mathbf{x}) = -\widehat{\mu}(\mathbf{x}) + c\,\widehat{\sigma}(\mathbf{x}), \tag{10}$$

where $c$ takes the following values, depending on the risk measure considered:

$$\begin{aligned}
c &= \mathrm{VaR}_{\alpha}(F_\star) = -q_\alpha^+(F_\star) \text{ is (minus) the } \alpha\text{-quantile of the } N(0,1) \text{ distribution} \\
c &= \mathrm{CVaR}_{\alpha}(F_\star) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_{u}(F_\star)\,du = \frac{\exp\{-\mathrm{VaR}_{\alpha}(F_\star)^2/2\}}{\alpha\sqrt{2\pi}} \\
c &= \rho_\phi(F_\star) = \int_0^1 \mathrm{VaR}_{u}(F_\star)\,\phi(u)\,du.
\end{aligned}$$

**Example 3.5** (ML risk estimators for the Laplace family)**.**
Consider now the Laplace with parameters $\boldsymbol{\theta} = (\mu, \lambda) \in \Theta = \mathbb{R} \times \mathbb{R}_+$. The ML estimators $\widehat{\theta} = (\widehat{\mu}, \widehat{\lambda})$ are

$$\widehat{\mu}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n x_i, \quad \widehat{s}(\mathbf{x}) = \widehat{\lambda}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n |x_i - \widehat{\mu}(\mathbf{x})|. \tag{11}$$

The resulting Laplace ML risk estimators are therefore:

$$\widehat{\rho}(\mathbf{x}) = -\widehat{\mu}(\mathbf{x}) + c\,\widehat{\lambda}(\mathbf{x}), \tag{12}$$

where $c$ takes the following values, depending on the risk measure considered:

$$\begin{aligned}
c &= \mathrm{VaR}_{\alpha}(F_\star) = -\log(2\alpha) \text{ for } \alpha \in (0, 1/2]) \\
c &= \mathrm{CVaR}_{\alpha}(F_\star) = -\frac{1}{\alpha}\int_0^\alpha \log(2u)\,du = -\log(2\alpha - 1) \text{ for } \alpha \in (0, 1/2]) \\
c &= \rho_\phi(F_\star) = -\int_0^{1/2} \log(2u)\,\phi(u)\,du + \int_{1/2}^1 \log(2 - 2u)\,\phi(u)\,du.
\end{aligned}$$

### 3.4 Effective risk measures

Let $\rho : \mathcal{D}_\rho \to \mathbb{R}$ be a risk measure and $\widehat{\rho} : \mathcal{X} \to \mathbb{R}$ a risk estimator. For a sample size $n \geq 1$ and any $F \in \mathcal{D}$, let $\mathcal{L}_n(\widehat{\rho}, F) \in \mathcal{D}$ be the law of the risk estimator

$$\mathcal{L}_n(\widehat{\rho}, F) = \mathrm{Law}(\widehat{\rho}(X_1, \ldots, X_n)), \quad \text{for } X_1, \ldots, X_n \sim F \text{ iid}.$$

A sensible requirement for a risk estimator is consistency:

**Definition 4** (Consistency)**.**
A risk estimator $\widehat{\rho}$ is (*weakly*) *consistent* with $\rho$ at $F \in \mathcal{D}_\rho$ if

$$\mathcal{L}_n(\widehat{\rho}, F) \to \delta_{\rho(F)} \quad \text{as } n \to \infty,$$

where $\delta_x$ denotes the point mass concentrated at $x \in \mathbb{R}$.

Given a risk measure $\rho$ that is continuous at some $F \in \mathcal{D}_\rho$, as a simple consequence of the Glivenko-Cantelli Theorem, we can see that the historical risk estimator $\widehat{\rho}^h$ is consistent with $\rho$ at $F$. For instance, we know that $\rho = \mathrm{VaR}_\alpha$ is continuous ar $F \in \mathcal{D}$ if and only if $q_\alpha^-(F) = q_\alpha^+(F)$. It follows that the historical estimator of $\mathrm{VaR}_\alpha$ is consistent with $\mathrm{VaR}_\alpha$ at any such $F$. On the other hand, nothing can be said a priori about consistency at those $F$ where $\rho$ is not continuous.

Consider now a consistent risk estimator $\widehat{\rho}(x_1, .., x_n)$ computed from an IID sample $X_1, ...X_n$. A key observation is that, since the estimator should be invariant under permutation of the data, its properties should only depend on the sample distribution $F^{\mathbf{x}}$. We can therefore define a map $\rho_{\text{eff}} : \mathcal{D}_{\text{emp}} \to \mathbb{R}$ such that

$$\widehat{\rho}(x_1, .., x_n) = \rho_{\text{eff}}(F^{\mathbf{x}}) \tag{13}$$

The map $\rho_{\text{eff}}$ associates a number to each (empirical) distribution and can thus be seen as a risk measure itself. The two-step risk estimation procedure can thus be represented as a *one-step* procedure in which the *effective risk measure* $\widehat{\rho}$ is applied to the empirical (i.e. historical) loss distribution $F^{\mathbf{x}}$. The above definition relation defines the effective risk measure $\rho_{\text{eff}}(F)$ for empirical distributions $F \in \mathcal{D}_{\text{emp}}$. This definition can be extended to a larger set of loss distributions using consistency:

**Definition 5** (Effective risk measure)**.** The *effective risk measure* associated to a risk measurement procedure with risk estimator $\widehat{\rho} : \mathcal{X} \to \mathbb{R}$ is

$$\widehat{\rho}(x_1, .., x_n) = \rho_{\text{eff}}(F^{\mathbf{x}}) \tag{14}$$

If $\widehat{\rho}$ is a consistent risk estimator, the effective risk measure admits a unique consistent extension to

$$\mathcal{D}_{\text{eff}} = \{F \in \mathcal{D} \,:\, \mathcal{L}_n(\widehat{\rho}, F) \to \delta_c \text{ for some } c \in \mathbb{R}\} \tag{15}$$

$$\text{by} \qquad \rho_{\text{eff}}(F) = \lim_{n\to\infty} \widehat{\rho}(X_1, .., X_n) \tag{16}$$

where $(X_i)_{i \geq 1}$ is any IID sequence with law $F$.

We note that by construction the risk estimator $\widehat{\rho}$ is consistent with $\rho_{\text{eff}}$ at all $F \in \mathcal{D}_{\text{eff}}$.

For a historical risk estimator $\widehat{\rho}^h$ which derives from $\rho$, the set of distributions for which $\widehat{\rho}$ is consistent with $\rho$ is exactly $\mathcal{D}_{\text{eff}}$ so for all $F \in \mathcal{D}_{\text{eff}}$, $\rho_{\text{eff}}(F) = \rho(F)$ and the effective risk measure coincides with the initial risk measure $\rho$.

But for parametric risk estimators, the effective risk measure associated can be very different from the risk measure initially considered. In the following we give some examples of effective risk measures associated to common parametric estimators.

**Example 3.6** (Gaussian loss distribution)**.** Consider the risk estimators introduced in Example 3.4. The associated effective risk measure is defined on $\mathcal{D}_2$ and given by

$$\rho_{\text{eff}}(F) = -\mu(F) + c\,\sigma(F), \tag{17}$$

where $\mu(F) = \int xF(dx)$ and $\sigma(F) = \sqrt{\int (x - \int xF(dx))^2 F(dx)}$.

**Example 3.7** (Laplace loss distribution)**.**
Consider the risk estimators introduced in Example 3.5. The associated effective risk measure is defined on $\mathcal{D}_1$ and given by

$$\rho_{\text{eff}}(F) = -\mu(F) + c\,\lambda(F), \tag{18}$$

where $\mu(F) = \int xF(dx)$ and $\lambda(F) = \int |x - \int xF(dx)|\, F(dx)$.

# 4 Qualitative robustness of risk estimators

If we take into account the errors in the estimation of the loss distribution of the portfolio, an important question is the impact of these errors on the risk measure of the portfolio. A risk estimator is said to be *robust* if a small variation in the loss distribution of the portfolio results in a small change in the distribution of the risk estimator.

## 4.1 $\mathcal{C}$-robustness of a risk estimator

In order to make this definition precise, one needs to quantify the notion of "small change" in the loss distribution and in the law of the estimator. We use here the Prohorov distance $d_P$ (see paragraph 2.1) to quantify the closeness of probability measures. We will denote $B_\delta(F)$ the ball centered in $F$ with radius $\delta$. Fix a set $\mathcal{C} \subseteq \mathcal{D}$ of plausible loss distributions and $F \in \mathcal{C}$. We assume $F$ is not an isolated point of $\mathcal{C}$, i.e. for any $\delta > 0$, there exists $G \neq F$ such that $G \in \mathcal{C}$ and $d_P(G, F) < \delta$. The intuitive notion of robustness can now be made precise using the following definition.

**Definition 6** ($\mathcal{C}$-robustness of a risk estimator)**.**
A risk estimator $\widehat{\rho}$ is *$\mathcal{C}$-robust* at $F$ if for any $\varepsilon > 0$ there exist $\delta > 0$ and $n_0 \geq 1$ such that:

$$G \in \mathcal{C},\ d_P(F, G) \leq \delta \quad \Longrightarrow \quad d_P(\mathcal{L}_n(\widehat{\rho}, F), \mathcal{L}_n(\widehat{\rho}, G)) \leq \varepsilon, \ \ \forall n \geq n_0.$$

Alternatively, the Levy metric can be used instead of the Prohorov metric. When $\mathcal{C} = \mathcal{D}$, i.e. when any perturbation of the loss distribution is allowed, the previous definition corresponds to the notion of *qualitative robustness* as proposed by Huber [12]. However, this case is not generally interesting in econometric or financial applications since requiring robustness against all perturbations of the model $F$ is quite restrictive.

In our case, the notion of $\mathcal{C}$-robustness depends on the class $\mathcal{C}$ of admissible distributions: for example the sample mean, which is well known not to be robust in the sense of Huber [12], is $\mathcal{C}$-robust at any $F \in \mathcal{D}_1$ if $\mathcal{C}$ is a set of distributions with support included in $]F^{-1}(\alpha), F^{-1}(1-\alpha)[$ for some $0 < \alpha < 1$.

Note that this notion of $\mathcal{C}$–robustness is *local* in the following sense. If $\widehat{\rho}$ is $\mathcal{C}$-robust at $F$ and $\mathcal{C}'$ is another set such that $\mathcal{C}' \cap B_\delta(F) = \mathcal{C} \cap B_\delta(F)$ for some $\delta > 0$, then $\widehat{\rho}$ is also $\mathcal{C}'$-robust.

### 4.2 Qualitative robustness of historical risk estimators

In this paragraph we focus on historical estimators $\widehat{\rho}^h(\mathbf{x}) = \rho(F^{\mathbf{x}})$, and study which among them are $\mathcal{C}$-robust, in the sense of Definition 6, where $\mathcal{C}$ is a suitable set of plausible loss distributions. The following proposition characterizes robustness among historical risk measures; a related, but slightly different result is due to Hampel [11].

**Proposition 2.** *Let $\rho$ be a risk measure, $\mathcal{C} \subseteq \mathcal{D}$ and $F \in \mathcal{C}$. If $\widehat{\rho}^h$ is consistent with $\rho$ at every $G \in \mathcal{C}$, the following are equivalent:*

1. *the restriction of $\rho$ to $\mathcal{C}$ is weakly continuous at $F$;*

2. *$\widehat{\rho}^h$ is $\mathcal{C}$-robust at $F$.*

*Proof.* First, observe that for any fixed $\varepsilon > 0$ and $G \in \mathcal{C}$, as $\widehat{\rho}^h$ is consistent with $\rho$ at $F$ and $G$, there exists $n^* \geq 1$ such that

$$d_P(\mathcal{L}_n(\widehat{\rho}^h, F), \delta_{\rho(F)}) + d_P(\mathcal{L}_n(\widehat{\rho}^h, G), \delta_{\rho(G)}) < \frac{2\varepsilon}{3}, \quad \forall n \geq n^*. \tag{19}$$

"1. $\Rightarrow$ 2". Assume that $\rho|_{\mathcal{C}}$ is continuous at $F$ and fix $\varepsilon > 0$. Then there exists $\delta > 0$ such that if $d_P(F, G) < \delta$, then $d_P(\delta_{\rho(F)}, \delta_{\rho(G)}) = |\rho(F) - \rho(G)| < \varepsilon/3$. Thus $\mathcal{C}$-robustness readily follows from (19) and the triangular inequality

$$\begin{aligned} d_P(\mathcal{L}_n(\widehat{\rho}^h, F), &\mathcal{L}_n(\widehat{\rho}^h, G)) \leq \\ &\leq d_P(\mathcal{L}_n(\widehat{\rho}^h, F), \delta_{\rho(F)}) + d_P(\delta_{\rho(F)}, \delta_{\rho(G)}) + \leq d_P(\mathcal{L}_n(\widehat{\rho}^h, G), \delta_{\rho(G)}). \end{aligned}$$

"2. $\Rightarrow$ 1". Conversely, assume that $\widehat{\rho}^h$ is $\mathcal{C}$-robust at $F$ and fix $\varepsilon > 0$. Then there exists $\delta > 0$ and $\overline{n} \geq 1$ such that

$$d_P(F, G) < \delta,\ G \in \mathcal{C} \quad \Rightarrow \quad d_P(\mathcal{L}_{\overline{n}}(\widehat{\rho}^h, F), \mathcal{L}_{\overline{n}}(\widehat{\rho}^h, G)) < \varepsilon/3.$$

As a consequence, from (19) and the triangular inequality

$$\begin{aligned} |\rho(F) - \rho(G)| &= d_P(\delta_{\rho(F)}, \delta_{\rho(G)}) \leq \\ &\leq d_P(\delta_{\rho(F)}, \mathcal{L}_n(\widehat{\rho}^h, F)) + d_P(\mathcal{L}_n(\widehat{\rho}^h, F), \mathcal{L}_n(\widehat{\rho}^h, G)) + d_P(\mathcal{L}_n(\widehat{\rho}^h, G), \delta_{\rho(G)}), \end{aligned}$$

it follows that $\rho|_{\mathcal{C}}$ is continuous at $F$. □

In view of the fact that a weakly continuous risk measure $\rho$ has a consistent historical risk estimator we have:

**Corollary 1.** *If $\rho$ is weakly continuous at any $G \in \mathcal{C}$, then $\widehat{\rho}^h$ is $\mathcal{C}$-robust at any $F \in \mathcal{C}$.*

Our analysis will use the following important result [12, Theorem 3.1]:

**Theorem 1.** *Let $\rho_m$ be a risk measure of the form (1), where $m$ is a probability measure on $(0, 1)$. Let $\overline{\alpha}$ be the largest $\alpha \geq 0$ such that supp$(m) \subseteq [\alpha, 1 - \alpha]$ and $\{\alpha_n\}_{n \geq 1}$ the (countable) set of point masses of $m$ (i.e. $m(\{\alpha_n\}) > 0$). Then:*

1. *if $\overline{\alpha} > 0$ then $\rho_m$ is weakly continuous at any $F \in \mathcal{D}_\rho$ such that $q_F^+(\alpha_n) = q_F^-(\alpha_n)$ for any $n$*
2. *if $\overline{\alpha} = 0$ then $\rho_m$ is not continuous at any $F \in \mathcal{D}_\rho$*

#### 4.2.1 Historical VaR$_\alpha$

Using the previous two results, we now show that the historical VaR is $\mathcal{C}$-robust, where $\mathcal{C}$ is the set of all distributions continuous at $\alpha$:

**Proposition 3** (Historical risk estimator of VaR$_\alpha$)**.**
*For $\alpha \in (0,1)$, let*

$$\mathcal{C}_\alpha \triangleq \{F \in \mathcal{D} \,:\, q_F^+(\alpha) = q_F^-(\alpha)\}.$$

*Then the historical VaR$_\alpha$ is $\mathcal{C}_\alpha$-robust at any $F \in \mathcal{C}_\alpha$.*

*Proof.* By using Theorem 1, we know that VaR$_\alpha$ is weakly continuous at any $F \in \mathcal{C}_\alpha$. Therefore, by applying Corollary 1 we obtain that the historical VaR$_\alpha$ is $\mathcal{C}_\alpha$-robust at any $F \in \mathcal{C}_\alpha$. □

This Proposition confirms the intuition that if the quantile of a distribution is uniquely determined, then the empirical quantile is a robust estimator.

#### 4.2.2 Historical CVaR and spectral risk measures

Let us now give a characterization of robustness when $m(du) = \phi(u)\,du$ for some density $\phi$ on $(0,1)$. This case will cover spectral risk measures and in particular expected shortfall/CVaR. For $p \in [1,\infty)$, we denote $\mathcal{D}_p$ the set of distributions with finite $p$-th moment.

**Proposition 4.** *Let $\phi$ be a density in $L^q(0,1)$, with $q \in (1,\infty]$, $1/p + 1/q = 1$. Let $\rho_\phi$ be the risk measure defined on $\mathcal{D}_p$ by*

$$\rho_\phi(F) = \int_0^1 VaR_u(F)\phi(u)\,du.$$

*Pick $F \in \mathcal{D}_p$, such that no discontinuity of $\phi$ coincides with a discontinuity of $q_F$, the quantile function of $F$. Then the historical estimator of $\rho_\phi$ is $\mathcal{D}_p$-robust at $F$ if and only*

$$supp(\phi) \subseteq [\overline{\alpha}, 1 - \overline{\alpha}]$$

*for some $\overline{\alpha} > 0$.*

*Proof.* Consider $F \in \mathcal{D}_p$, such that no discontinuity of $\phi$ coincides with a discontinuity of $q_F$. By contraposition, we assume that $\overline{\alpha} = 0$. Then by using Theorem 1 we know that $\rho_\phi$ is not continuous at $F$. As $\rho_\phi$ is defined

on $\mathcal{D}_p$, it is equivalent to saying that the restriction of $\rho_\phi$ to $\mathcal{D}_p$ is not continuous at $F$. A result of Van Zwet [18] implies that the historical estimator of $\rho_\phi$ is consistent with $\rho_\phi$ at any $G \in \mathcal{D}_p$. Thus by applying Proposition 2 we conclude that the historical estimator of $\rho_\phi$ is not $\mathcal{D}_p$-robust at $F$.
Now we assume that $\overline{\alpha} > 0$, then Theorem 1 shows that $\rho_\phi$ is weakly continuous at $F$, which is equivalent to saying that the restriction of $\rho_\phi$ to $\mathcal{D}_p$ is weakly continuous at $F$. Therefore, by using Proposition 2, we see that the historical estimator of $\rho_\phi$ is $\mathcal{D}_p$-robust at $F$. □

**Corollary 2.** *Proposition 4 implies that:*

1. *The historical $CVaR_\alpha$ is not robust on $\mathcal{D}_1 = \{F \in \mathcal{P}, \int_{-\infty}^{+\infty} |x|\, dF < \infty\}$. Indeed, in this case $\phi(u) = \frac{1_{[0,\alpha]}}{\alpha}$ and supp$(\phi) = [0, \alpha]$.*
2. *More generally, historical estimators of spectral risk measures, i.e. with a **decreasing function** $\phi$ in $L^q(0,1)$, are not $\mathcal{D}_p$-robust at any $F \in \mathcal{D}_p$ for $1/p + 1/q = 1$.*

This proposition stresses a conflict between subadditivity and robustness, which means that requiring coherence to a risk measure contrasts with the robustness of its historical estimator.

More generally, a historical estimator of any coherent risk measure defined as the supremum of a finite number of coherent spectral risk measures:

$$\rho_\phi(F) = \sup_{i=1,..,K} \int_0^1 \text{VaR}_\alpha(F)\, \phi_i(\alpha)\, d\alpha \tag{20}$$

with **decreasing functions** $\phi_i \geq 0$ **in** $L^q(0,1)$ **for** $i = 1,..,K$, is not $\mathcal{D}_p$-robust at any $F \in \mathcal{D}_p$.

Kusuoka [15] has shown that any coherent distribution-based risk measure can be represented in the form above where the supremum is taken over a possibly infinite family of decreasing weight functions $\phi$; so the above statement has a fairly general flavor.

#### 4.2.3 A robust tail risk measure

From Proposition 1, the sub-additivity of $\rho_\phi$ implies that $\phi$ is decreasing while the robustness of the historical estimator $\widehat{\rho}^h$ implies that supp$(\phi) \subset [\alpha, 1-\alpha]$ for some $\alpha > 0$. These two requirements are clearly not compatible, which reveals a conflict between robustness and subadditivity.

Although historical estimators of $\mathrm{CVaR}_{\alpha}$ are not robust, one can easily modify them to obtain a robust weighted VaR, defined for $F \in \mathcal{D}_1$ by

$$WVaR_{\alpha_1,\alpha_2}(F) = \frac{1}{\alpha_2 - \alpha_1} \int_{\alpha_1}^{\alpha_2} d\alpha \mathrm{VaR}_\alpha(F) \tag{21}$$

**Corollary 3.** *Let us consider $0 < \alpha_1 < \alpha_2 < 1$ and a distribution $F$ in $\mathcal{D}_1$ such that $q_F^+(\alpha_1) = q_F^-(\alpha_1)$ and $q_F^+(\alpha_2) = q_F^-(\alpha_2)$. Then the historical estimator of $WVaR_{\alpha_1,\alpha_2}$ is $\mathcal{D}_1$-robust at $F$.*

Due to the conflict between subadditivity and robustness noted above, $WVaR_{\alpha_1,\alpha_2}$ is not (asymptotically) sub-additive, but for small $\alpha_1$ it is in fact empirically undistinguishable from $\mathrm{CVaR}_{\alpha_2}$. This point will be further discussed in Section 7.

### 4.3 Qualitative robustness of parametric risk estimators

We now discuss robustness properties of parametric risk estimators of translation invariant and homogeneous risk measures (for instance any risk measure $\rho_m$ of type (1)). Pick a distribution $F_\star \in \mathcal{D}$ and fix the location parameter to $\mu_0$. We consider in this section scale parameters that are computed via the Maximum Likelihood (ML). For $F \in \mathcal{D}_{F_\star,\mu_0}$ we denote by $s(F)$ the scale parameter of $F$. Define

$$\psi_\star(x) = -1 - x\frac{f_\star'(x)}{f_\star(x)} \tag{22}$$

Then the scale parameter is a solution of

$$\lambda(s, F) \triangleq \int \psi_\star\left(\frac{x-\mu_0}{s}\right) F(dx) = 0 \quad \text{for} \quad F \in \mathcal{D}_{F_\star,\mu_0} \tag{23}$$

By defining $\mathcal{D}_{\psi_\star} = \{F \in \mathcal{D} | \int \psi_\star(x)F(dx) < \infty\}$, we can define $s(F)$ for any $F \in \mathcal{D}_{\psi_\star}$. Note that if $F \notin \mathcal{D}_{F_\star,\mu_0}$, $s(F)$ does not correspond to the "scale parameter" of $F$. Moreover if we compute the scale parameter of $F_{\mathbf{x}} \in \mathcal{D}_{\mathrm{emp}}$ we recover the classical ML estimator $\widehat{s}(\mathbf{x})$ presented in section 3.

**Example 4.1** (Gaussian family)**.** The function $\psi_\star^g$ for the Gaussian family can be easily computed

$$\psi_\star^g(x) = -1 + x^2, \tag{24}$$

and we get $\mathcal{D}_{\psi_\star^g} = \mathcal{D}_2$ and recover the well-known MLE of the Gaussian variance

$$\widehat{s}(\mathbf{x}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2}. \tag{25}$$

**Example 4.2** (Laplace family)**.** The scale parameter of the Laplace family corresponds to $\lambda$. The function $\psi^l_\star$ is the following

$$\psi^l_\star(x) = -1 + |x|, \tag{26}$$

and we get $\mathcal{D}_{\psi^l_\star} = \mathcal{D}_1$ and obtain the following ML estimator of $s$

$$\widehat{s}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \mu_0|. \tag{27}$$

The following result exhibits conditions on the function $\psi_\star$ under which the maximum likelihood estimator of the scale parameter is weakly continuous on $\mathcal{D}_{\psi_\star}$:

**Theorem 2** (Weak continuity of the MLE)**.** *Let us consider $\mathcal{D}_{\psi_\star}$, a scale family associated to $\psi_\star$ and $s : \mathcal{D}_{\psi_\star} \to \mathbb{R}^+$ the scale function defined on $\mathcal{D}_{\psi_\star}$. Suppose now that $\psi_\star$ is even, monotone and increasing on $\mathbb{R}^+$ and takes values of both signs. Then, these two assertions are equivalent*

- *$s$ is weakly continuous at $F \in \mathcal{D}_{\psi_\star}$*
- *$\psi_\star$ is bounded and $\lambda(s, F) \triangleq \int \psi_\star\left(\frac{x-\mu_0}{s}\right) F(dx) = 0$ has a unique solution $s = s(F)$ for $F \in \mathcal{D}_{\psi_\star}$.*

*Proof.* Consider $\mathcal{D}_{\psi_\star}$ a scale family associated to $\psi_\star$. For simplicity, we consider here that $\mu_0 = 0$. We can assume that $F \in \mathcal{D}_{\psi_\star}$ is the distribution of the random variable $X$ representing the profits and losses of a portfolio. We will show that the continuity problem of the scale function $s : \mathcal{D}_{\psi_\star} \to \mathbb{R}^+$ of portfolios $X$ can be reduced to the continuity (on a properly defined space) issue of the location function of portfolios $Y = \log(X^2)$. The change of parameter here is made to use results of [12] concerning weak continuity of location parameters.

By denoting $G$ the distribution of $Y$, we get $G(y) = P(Y < y) = P(X^2 < e^y) = F(e^{y/2}) - F(-e^{y/2})$, with density $g(y) = G'(y) = e^{y/2}\, f(e^{y/2})$. Moreover, if we denote $\tau$ the location function defined on the location family of distributions $\mathcal{D}_{\varphi_\star} \triangleq \{G : \int \varphi_\star(y) G(dy) < \infty\}$, where $\varphi_\star(y-\tau) = -\frac{g'_\star(y-\tau)}{g_\star(y-\tau)}$, then it is defined by the solution of the following implicit relation

$$\int \varphi_\star(y-\tau) G(dy) = 0 \tag{28}$$

for $G \in \mathcal{D}_{\varphi_\star}$. Let us now replace $g_\star$ by its value w.r.t $f_\star$

$$\begin{aligned}
\varphi_\star(y-\tau) &= -\frac{g'_\star(y-\tau)}{g_\star(y-\tau)} \\
&= -\frac{\frac{1}{2}e^{(y-\tau)/2}f_\star(e^{(y-\tau)/2}) + \frac{1}{2}e^{y-\tau}f'_\star(e^{(y-\tau)/2})}{e^{(y-\tau)/2}f_\star(e^{(y-\tau)/2})} \\
&= -\frac{1}{2}\left[1 + e^{(y-\tau)/2}\ \frac{f'_\star(e^{(y-\tau)/2})}{f_\star(e^{(y-\tau)/2})}\right] \\
&= -\frac{1}{2}\left[1 + \frac{x}{s}\ \frac{f'_\star(\frac{x}{s})}{f_\star(\frac{x}{s})}\right] \\
&= \frac{1}{2}\psi_\star\left(\frac{x}{s}\right). \qquad (29)
\end{aligned}$$

Moreover, we can also rewrite $G(dy)$ as

$$G(dy) = g'(y)dy = xf(x)d\left(\log(x^2)\right) = 2f(x)dx = 2F(dx),$$

which ends to show the equivalence between the function $s(F)$ for $F \in \mathcal{D}_{\psi_\star}$ and the function $\tau(G)$ for $G \in \mathcal{D}_{\varphi_\star}$. We have shown that a scale function characterized by the function $\psi_\star$, can also be interpreted as a location function characterized by the function $\varphi_\star$. Therefore, the weak continuity of the scale function will depend on the behavior of $\varphi_\star$. From Equation (29), we see that for all $x \in \mathbb{R}$, $2\,\varphi_\star(x) = \psi_\star[e^{x/2}]$. Therefore, as the associated function $\psi_\star$ to the function $s(.)$ is assumed to be even, and monotone increasing on $\mathbb{R}^+$, it implies that $\varphi_\star$ is monotone increasing on $\mathbb{R}$. Moreover, as $\psi_\star$ takes values of both signs it is also true for $\varphi_\star$. To conclude, we apply [12, Theorem 2.6] which states that a location function associated to $\varphi_\star$ is weakly continuous at $G$ if and only if $\varphi_\star$ is bounded and the location function computed at $G$ is unique. □

Theorem 2 enables us to study the robustness of parametric risk estimators for Gaussian or Laplace scale families i.e. with a fixed location parameter $\mu = \mu_0$:

**Corollary 4** (Parametric risk estimators for Gaussian and Laplace scale families)**.** *Gaussian (resp. Laplace) risk estimators of translation invariant and homogeneous risk measures are not $\mathcal{D}_2$-robust (resp. $\mathcal{D}_1$-robust) at any $F$ in $\mathcal{D}_2$ (resp. in $\mathcal{D}_2$).*

*Proof.* We detail the proof for the Gaussian scale family. The same argument can be used for the Laplace scale family as well. Let us consider a Gaussian risk estimator of a translation invariant and homogeneous risk measure, denoted $\widehat{\rho}(\mathbf{x}) = -\mu_0 + c\,\widehat{\sigma}(\mathbf{x})$. First of all we notice that the function $\psi_\star^g$ associated to the ML estimation of the scale parameter of a distribution belonging to the Gaussian scale family is even, and increasing on $\mathbb{R}^+$. Moreover it takes values of both signs. Secondly, we recall that the effective risk measure associated to the Gaussian risk estimator is $\rho_{\text{eff}}(F) = -\mu_0 + c\,\sigma(F)$ for all $F \in \mathcal{D}_{\text{eff}} = \mathcal{D}_{\psi_\star^g} = \mathcal{D}_2$. Therefore, as $\psi_\star^g$ is unbounded, by using Theorem 2, we know that $\rho_{\text{eff}}$ is not continuous at any $F \in \mathcal{D}_2$. As the Gaussian risk estimator considered $\widehat{\rho}$ verifies $\widehat{\rho}(\mathbf{x}) = \widehat{\rho_{\text{eff}}}^h(\mathbf{x})$, and is consistent with $\rho_{\text{eff}}$ at all $F \in \mathcal{D}_2$ by construction, we can apply Proposition 2 to conclude that, for $F \in \mathcal{D}_2$, $\widehat{\rho}$ is not $\mathcal{D}_2$-robust at $F$. $\square$

# 5 Sensitivity analysis of risk measurement procedures

The discussion above leads to distinguish robust from non-robust risk measurement procedures. This qualitative analysis can be further refined through a quantitative sensitivity analysis that we now discuss. Intuitively, the empirical sensitivity of a risk estimator, computed from an IID sample of size $N$ with law $F$, to the addition of a new observation can be measured by

$$
\begin{aligned}
S_N(x,F) &= \frac{\widehat{\rho}(X_1,...,X_N,X_{N+1}) - \widehat{\rho}(X_1,...,X_N)}{\frac{1}{N+1}} \\
&= \frac{\rho_{\text{eff}}(\frac{1}{N+1}\sum_{i=1}^N 1_{u\geq X_i} + 1_{u\geq x}) - \rho_{\text{eff}}(\frac{1}{N}\sum_{i=1}^N 1_{u\geq X_i})}{\frac{1}{N+1}}
\end{aligned}
$$

We will call this quantity the empirical sensitivity function of the risk estimator $\widehat{\rho}$. This motivates the following definition in the large sample case:

**Definition 7** (Sensitivity function of a risk measurement procedure)**.** The *sensitivity function* of a risk measurement procedure $(M,\rho)$ at $F \in \mathcal{D}_{\text{eff}}$ is the real-valued map $S$ defined by

$$S(z;F) \triangleq \lim_{\varepsilon\to 0^+} \frac{\rho_{\text{eff}}(\varepsilon\delta_z + (1-\varepsilon)F) - \rho_{\text{eff}}(F)}{\varepsilon}$$

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

$S(z,F)$ measures the sensitivity of the risk estimator based on a large sample to the addition of a new data point. $S(z;F)$ is nothing but the directional derivative of the effective risk measure $\rho_{\text{eff}}$ at $F$ in the direction $\delta_z \in \mathcal{D}$. In the language of robust statistics, it is the influence function of the risk estimator $\widehat{\rho}$ and is related to the asymptotic variance of the historical estimator of $\rho$ [12, 11].

**Remark 1.** If $\mathcal{D}_\rho$ is convex and contains all empirical distributions, then $\varepsilon\delta_z + (1-\varepsilon)F \in \mathcal{D}_\rho$ for any $\varepsilon \in [0,1]$, $z \in \mathbb{R}$ and $F \in \mathcal{D}_\rho$. These conditions hold for all the risk measures we are considering.

## 5.1 Historical VaR

We have seen before that the effective risk measure associated to $\widehat{\mathrm{VaR}}^h_\alpha$ is the restriction of $\mathrm{VaR}_\alpha$ to

$$\mathcal{D}_{\text{eff}} = \mathcal{C}_\alpha = \{F \in \mathcal{D} \,:\, q^-_\alpha(F) = q^+_\alpha(F)\}.$$

We want to compute the sensitivity function of the historical $\mathrm{VaR}_\alpha$, at least in the plain case when $F \in \mathcal{D}$ admits a strictly positive density $f$; it is immediate to see that in such a case $F \in \mathcal{C}_u$ for any $u \in (0,1)$.

**Proposition 5.** *If $F \in \mathcal{D}$ admits a strictly positive density $f$, then the sensitivity function at $F$ of the historical $\mathrm{VaR}_\alpha$ is*

$$S(z) = \begin{cases} \dfrac{1-\alpha}{f(q_\alpha(F))} & \text{if } z < q_\alpha(F) \\ 0 & \text{if } z = q_\alpha(F) \\ -\dfrac{\alpha}{f(q_\alpha(F))} & \text{if } z > q_\alpha(F) \end{cases} \tag{30}$$

*Proof.* First we observe that the map $u \mapsto q(u) \triangleq q_u(F)$ is the inverse of $F$ and so it is differentiable at any $u \in (0,1)$ and we have:

$$q'(u) = \frac{1}{F'(q(u))} = \frac{1}{f(q_u(F))}.$$

Fix $z \in \mathbb{R}$ and set, for $\varepsilon \in [0,1)$, $F_\varepsilon = \varepsilon\delta_z + (1-\varepsilon)F$, so that $F \equiv F_0$. For $\varepsilon > 0$, the distribution $F_\varepsilon$ is differentiable at any $x \neq z$, with $F'_\varepsilon(x) = (1-\varepsilon)f(x) > 0$, and has a jump (of size $\varepsilon$) at the point $x = z$. Hence, for any $u \in (0,1)$ and $\varepsilon \in [0,1)$, $F_\varepsilon \in \mathcal{C}_u$, i.e. $q^-_u(F_\varepsilon) = q^+_u(F_\varepsilon) \triangleq q_u(F_\varepsilon)$. In particular we can easily compute:

$$q_\alpha(F_\varepsilon) = \begin{cases} q(\frac{\alpha}{1-\varepsilon}) & \text{for } \alpha < (1-\varepsilon)F(z) \\ q(\frac{\alpha-\varepsilon}{1-\varepsilon}) & \text{for } \alpha \geq (1-\varepsilon)F(z) + \varepsilon \\ z & \text{otherwise} \end{cases} \tag{31}$$

Assume now that $z > q(\alpha)$, i.e. $F(z) > \alpha$; from (31) it follows that

$$q_\alpha(F_\varepsilon) = q(\tfrac{\alpha}{1-\varepsilon}), \quad \text{for } \varepsilon < 1 - \tfrac{\alpha}{F(z)}.$$

As a consequence

$$\begin{aligned}
S(z) &= \lim_{\varepsilon \to 0^+} \frac{\mathrm{VaR}_\alpha(F_\varepsilon) - \mathrm{VaR}_\alpha(F_0)}{\varepsilon} = -\frac{d}{d\varepsilon} q_\alpha(F_\varepsilon)|_{\varepsilon=0} \\
&= -\frac{d}{d\varepsilon} q\left(\frac{\alpha}{1-\varepsilon}\right)\Big|_{\varepsilon=0} = -\left[\frac{1}{f(q(\frac{\alpha}{1-\varepsilon}))}\frac{\alpha}{(1-\varepsilon)^2}\right]_{\varepsilon=0} \\
&= -\frac{\alpha}{f(q_\alpha(F))}
\end{aligned}$$

The case $q(\alpha) < z$ is handled in a very similar way. Finally, if $z = q(\alpha)$, then, again by (31) we have $q_\alpha(F_\varepsilon) = z$ for any $\varepsilon \in [0, 1)$. Hence, in this case

$$S(z) = -\frac{d}{d\varepsilon} q_\alpha(F_\varepsilon)|_{\varepsilon=0} = 0,$$

and we conclude. □

This example shows that the historical $\mathrm{VaR}_\alpha$ has a bounded sensitivity to a change in the data set, which means that this risk measurement procedure is not very sensitive to a small change in the data set.

### 5.2 Historical estimators of spectral risk measures and CVaR

Consider a distribution $F$ having positive density $f > 0$. Assume that:

$$\int_0^1 \frac{\phi(u)}{f(q_u(F))}\, du < \infty.$$

**Proposition 6.** *The sensitivity function at $F \in \mathcal{D}_\phi$ of the historical estimation procedure of $\rho_\phi$ is*

$$S(z) = -\int_0^1 \frac{u}{f(q_u(F))}\,\phi(u)\,du + \int_{F(z)}^1 \frac{1}{f(q_u(F))}\,\phi(u)\,du$$

*Proof.* Using the notations of the previous proof we have:

$$\begin{aligned} S(z) &= \lim_{\varepsilon \to 0^+} \int_0^1 \frac{\mathrm{VaR}_u(F_\varepsilon) - \mathrm{VaR}_u(F)}{\varepsilon}\,\phi(u)\,du \\ &= \int_0^1 \lim_{\varepsilon \to 0^+} \frac{\mathrm{VaR}_u(F_\varepsilon) - \mathrm{VaR}_u(F)}{\varepsilon}\,\phi(u)\,du \\ &= \int_0^1 - \left[\frac{d}{d\varepsilon} q_u(F_\varepsilon)\right]_{\varepsilon=0} \phi(u)\,du \\ &= \int_0^{F(z)} \frac{-u}{f(q_u(F))}\,\phi(u)\,du + \int_{F(z)}^1 \frac{1-u}{f(q_u(F))}\,\phi(u)\,du, \end{aligned}$$

thanks to Proposition 5. We stress that changing the integral with the limit in the second equality above is legitimate. Indeed, $\lim_{\varepsilon\to 0^+} \varepsilon^{-1}(\mathrm{VaR}_u(F_\varepsilon)) - \mathrm{VaR}_u(F))$ exists, is finite for all $u \in (0,1)$, and for $\varepsilon$ small

$$\left| \frac{\mathrm{VaR}_u(F_\varepsilon) - \mathrm{VaR}_u(F)}{\varepsilon} \right| < \frac{1}{f(q_u(F))} \in L^1(\phi),$$

so that we can apply dominated convergence. $\square$

Since the effective risk measure associated to historical $\mathrm{CVaR}_\alpha$ is $\mathrm{CVaR}_\alpha$ itself, defined on $\mathcal{D}^- = \{F \in \mathcal{D} : \int x^- F(dx) < \infty\}$, an immediate consequence of the previous proposition is the following

**Corollary 5.** *The sensitivity function at $F \in \mathcal{D}^-$ for historical $CVaR_\alpha$ is*

$$S(z) = \begin{cases} -\dfrac{z}{\alpha} + \dfrac{1-\alpha}{\alpha}\, q_\alpha(F) - \mathrm{CVaR}_\alpha(F) & \textit{if } z \le q_\alpha(F) \\[2ex] -q_\alpha(F) - \mathrm{CVaR}_\alpha(F) & \textit{if } z \ge q_\alpha(F) \end{cases}$$

This result shows that the sensitivity of historical $\mathrm{CVaR}_\alpha$ is linear in $z$, and thus unbounded. It means that this risk measurement procedure is less robust than the historical $\mathrm{VaR}_\alpha$.

### 5.3 Parametric estimators: Gaussian model

We have seen that the effective risk measure of the Gaussian estimators of VaR, CVaR, or any spectral risk measure is

$$\rho_{\text{eff}}(F) = -\mu(F) + c\sigma(F), \quad F \in \mathcal{D}_{\text{eff}} = \mathcal{D}_2,$$

where $c = \rho(N(0,1))$ is a constant depending only on the risk measure $\rho$ (we are not interested in its explicit value here).

**Proposition 7.** *The sensitivity function at $F \in \mathcal{D}_2$ of the Gaussian estimator associated to a translation invariant and homogeneous risk measure $\rho$ is*

$$S(z) = \mu - z + \frac{\sigma\, c}{2}\left[\left(\frac{z-\mu}{\sigma}\right)^2 - 1\right]$$

*where $c = \rho(N(0,1))$. For the Gaussian scale family ($\mu = \mu_0$) we have:*

$$S(z) = \frac{\sigma\, c}{2}\left[\left(\frac{z-\mu_0}{\sigma}\right)^2 - 1\right].$$

*Proof.* Let, for simplicity, $\mu = \mu(F)$ and $\sigma = \sigma(F)$. Fix $z \in \mathbb{R}$ and set, as usual, $F_\varepsilon = (1-\varepsilon)F + \varepsilon\delta_z$ ($\varepsilon \in [0,1)$); observe that $F_\varepsilon \in \mathcal{D}^2$ for any $\varepsilon$. If we set $\psi(\varepsilon) \triangleq -\mu(F_\varepsilon) + c\sigma(F_\varepsilon)$, with $c = \rho(N(0,1))$, then we have $S(z) = \psi'(0)$. It is immediate to compute $\mu(F_\varepsilon) = (1-\varepsilon)\mu + \varepsilon z$ and

$$\begin{aligned}
\sigma^2(F_\varepsilon) &= \int_{\mathbb{R}} x^2\, F_\varepsilon(dx) - \mu(F_\varepsilon)^2 \\
&= (1-\varepsilon)\int_{\mathbb{R}} x^2\, F(dx) + \varepsilon z^2 - [(1-\varepsilon)\mu + \varepsilon z]^2 \\
&= (1-\varepsilon)(\sigma^2+\mu^2) + \varepsilon z^2 - (1-\varepsilon)^2\mu^2 - \varepsilon^2 z^2 - 2\varepsilon(1-\varepsilon)\mu z \\
&= \sigma^2 + \varepsilon[(z-\mu)^2 - \sigma^2] - \varepsilon^2(\mu^2 + \sigma^2 - 2\mu z)
\end{aligned}$$

As a consequence

$$\begin{aligned}
\psi'(0) &= \frac{d}{d\varepsilon}\left[-(1-\varepsilon)\mu - \varepsilon z + c\sqrt{\sigma^2(F_\varepsilon)}\right]_{\varepsilon=0} \\
&= \mu - z + \frac{\sigma\, c}{2}\left[\left(\frac{z-\mu}{\sigma}\right)^2 - 1\right]
\end{aligned}$$

as we desired. Similarly, starting from $\psi(\varepsilon) = -\mu_0 + c\sigma(F_\varepsilon)$, we obtain the second statement. □

## 5.4 Parametric estimators: double-exponential model

Consider now a translation invariant and homogeneous risk measure $\rho$ and its MLE estimator under the Laplace family. We have seen that the effective risk measure associated to this risk estimator is

$$\rho_{\text{eff}}(F) = -\mu(F) + c\,\lambda(F), \quad F \in \mathcal{D}_{\text{eff}} = \mathcal{D}^1,$$

where $c = \rho(G)$, and $G$ is the distribution with density $g(x) = e^{-|x|}/2$.

| Risk measurement procedure | Dependence in $z$ of $S(z)$ |
| --- | --- |
| Historical VaR | bounded |
| Gaussian VaR | quadratic |
| Laplace VaR | linear |
| Historical CVaR | linear |
| Gaussian CVaR | quadratic |
| Laplace CVaR | linear |

Table 1: Behavior of sensitivity functions for some important examples of risk measurement procedures.

**Proposition 8.** *Let $\rho$ be a translation invariant and homogeneous risk measure. The sensitivity function at $F \in \mathcal{D}^1$ of its parametric estimator based on the Laplace scale family with location parameter $\mu = \mu_0$ is*

$$S(z) = \lambda\, c \left[ \frac{|z - \mu_0|}{\lambda} - 1 \right]$$

*where $\lambda = \lambda(F)$.*

*Proof.* As usual, we have, for $z \in \mathbb{R}$, $S(z) = \psi'(0)$, where $\psi(\varepsilon) = -\mu_0 + c\,\lambda(F_\varepsilon)$, $F_\varepsilon = (1-\varepsilon)F + \varepsilon\delta_z$ and $c$ is defined above. For simplicity, let us denote $\lambda = \lambda(F)$. We have

$$\psi(\varepsilon) = c\,(1-\varepsilon)\lambda + c\,\varepsilon\,|z - \mu_0|,$$

hence

$$\psi'(0) = c|z - \mu_0| - c\,\lambda.$$

□

This proposition shows that the sensitivity of the Laplace risk estimator at any $F \in \mathcal{D}_1$ is not bounded but linear in $z$. Nonetheless, the sensitivity of the Gaussian risk estimator is quadratic at any $F \in \mathcal{D}_2$, which indicates a higher sensitivity to outliers in the data set.

## 6 An example

We consider in this section an application of the above concepts to a measure of counterparty risk for a derivatives portfolio. The data for the example studied here has been provided to us by Société Générale Risk Management unit: it consists of 1000 risk scenarios for the bank portfolio, simulated using the internal model of the bank and incorporating hundreds of different risk factors (market risk, interest rate risk, counterparty default). Each scenario is simulated over 30 years and the objective is to quantify the counterparty risk exposure of the bank's derivatives portfolio via an appropriate tail risk measure of the loss distribution at various horizons (e.g. 1 day, 1 month, 1 year and 10 years). In our study, we have considered the 1 day horizon and plotted the portfolio histogram in figure 1.

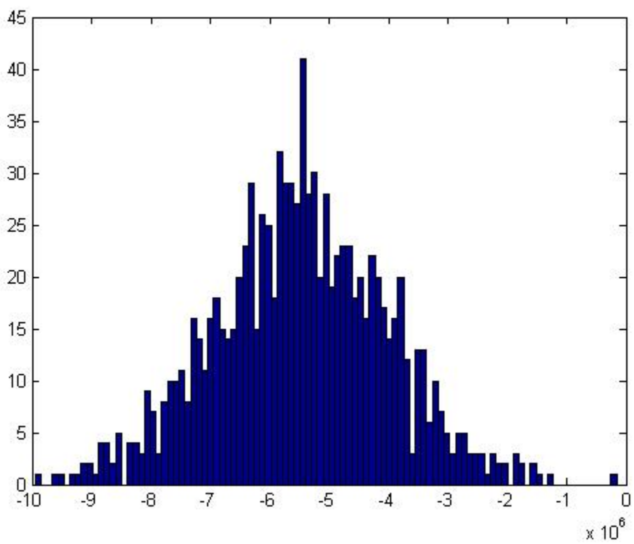



Figure 1: Histogram of portfolio gain at a 1 day horizon.

In table 2, we have computed for different quantile levels $\alpha = 1\%, 0.4\%$ the following risk estimators: historical, gaussian, and laplace VaR and historical, gaussian, and laplace CVaR.

Then, in order to analyze the accuracy of our theoretical risk estimator sensitivities, we have computed the empirical sensitivities of historical, gaussian, and laplace VaR and historical, gaussian, and laplace CVaR and plotted them with the theoretical one. The result are shown in figures 2 and 3, where the $x$-axis represents the value of the portfolio, and the $y$-axis the sensitivity to the corresponding value of the portfolio.

One can notice that the theoretical and empirical sensitivities coincide for all risk estimators except for historical risk measurement procedures. For the historical CVaR, the theoretical sensitivity is very close to the empirical

| Risk measurement procedure | $\alpha = 1\%$ | $\alpha = 0.4\%$ |
|---|---|---|
| Historical VaR | 8.887 | 9.193 |
| Historical CVaR | 9.291 | 9.606 |
| Gaussian VaR | 8.876 | 9.351 |
| Gaussian CVaR | 9.370 | 9.802 |
| Laplace VaR | 9.970 | 11.021 |
| Laplace CVaR | 11.117 | 12.167 |

Table 2: Risk estimators for $\alpha = 1\%, 0.4\%$ at a 1 day horizon computed for several risk measurement procedures (in million euro).

one. Nonetheless, we notice that the empirical sensitivity of the historical VaR can be equal to 0 because it is strongly dependent on the integer part of $N\ \alpha$, where $N$ is the number of scenarios and $\alpha$ the order of the quantile. This dependency disappears asymptotically.

While the empirical sensitivities, which require perturbating the data sets and recomputing the risk measures, are quite costly to compute, the sensitivity functions are analytically and thus easily computable. The excellent agreement for realistic sample sizes shown in these examples implies that our theoretical sensitivity functions are useful for evaluating the sensitivity of risk estimators in such practical settings.

# 7 Discussion

## 7.1 Summary of main results

Let us now summarize the contributions and main conclusions of this study.

First, we have argued that when the estimation step is explicitly taken into account in a risk measurement procedure, issues like robustness and sensitivity to the data set are important and need to be accounted for with at least the same attention as the coherence properties set forth by Artzner et al [3]. Indeed, an unstable/non-robust risk estimator, be it related to a coherent measure of risk, is of little use in practice.

Second, we have shown that the choice of the estimation method matters when discussing the robustness of risk measurement procedures: our examples show that different estimation methods coupled with the same risk measure lead to very different properties in terms of robustness and sensitivity.

Nonparametric ("historical") VaR is a qualitatively robust estimation procedure, whereas the proposed examples of coherent (distribution-based) risk measures do not pass the test of qualitative robustness. Most parametric estimation procedures for VaR and CVaR lead to nonrobust estimators. These observations plead against the use of CVaR. On the other hand historical VaR and more generally weighted averages of historical VaR of the type

$$\frac{1}{\alpha_2 - \alpha_1} \int_{\alpha_1}^{\alpha_2} \mathrm{VaR}_u(F)\, du$$

lead to robust empirical estimators.

### 7.2 Re-examining subadditivity

The conflict we have noted between robustness of a risk measurement procedure and the subadditivity of the risk measure shows that one cannot achieve robust estimation in this framework while preserving subadditivity. While a strict adherence to the coherence axioms of Artzner et al [3] would push us to choose subadditivity over robustness, several recent studies [7, 13, 14] have provided reasons for not doing so.

Danielsson et al. [7] explore the potential for violations of VaR subadditivity and report that for most practical applications VaR is sub-additive. They conclude that in practical situations there is no reason to choose a more complicated risk measure than VaR, solely for reasons of subadditivity. Arguing in a different direction, Ibragimov & Walden [13] show that for very "heavy-tailed" risks defined in a very general sense, diversification does not necessarily decrease tail risk but actually can increase it, in which case requiring sub-additivity would in fact be unnatural. Finally, Kou et al [14] argue against subadditivity from an *axiomatic* viewpoint and propose to replace it by a weaker property of *co-monotonic subadditivity.* All these objections to the sub-additivity axiom deserve serious consideration and further support the choice of robust risk measurement procedures over non-robust ones for the sole reason of saving sub-additivity.

Without taking an axiomatic stance on these issues, we hope to have convinced the reader that there is more to risk measurement than the choice of a "risk measure". We think that the property of robustness - and not only the coherence - should be a concern for regulators and end-users when designing risk measurement procedures. What our study illustrates is that the design of robust risk estimation procedures requires the inclusion of the statistical estimation step in the risk measurement procedure. We hope this work will stimulate further discussion on these important issues.

## References


[1] Acerbi C. (2002) *Spectral measures of risk : A coherent representation of subjective risk aversion.* Journal of Banking & Finance 26 1505-1518.

[2] Acerbi C. (2007) *Coherent measures of risk in everyday market practice.* Quantitative Finance, Vol. 7, No. 4, August 2007, 359–364.

[3] Artzner P., Delbaen F., Eber J.M., Heath D. (1999) *Coherent measures of risk.* Mathematical Finance, 9, No. 3, 203-228.

[4] Cont R. (2006) *Model uncertainty and its impact on the pricing of derivative instruments.* Mathematical Finance, 16, No. 3, 519-547.

[5] Costa V., Deshayes J. (1977) *Comparaison des R-L-M estimateurs.* Théorie de la robustesse et estimation d'un paramètre, Astérisque 43-44.

[6] Deniau C., Oppenheim G., Viano C. (1977) *Courbe d'influence et sensibilité.* Théorie de la robustesse et estimation d'un paramètre, Astérisque 43-44.

[7] Danielsson, J., Jorgensen, B. N., Sarma, M., Samorodnitsky, G., and de Vries, C. G. (2005). Subadditivity re-examined: the case for Value at risk, Working Paper.

[8] Föllmer H., Schied A. (2002) *Convex measures of risk and trading constraints*, Finance and Stochastics, 6, No. 4, 429-447.

[9] Föllmer H., Schied A. (2004) *Stochastic Finance.*, Studies in Mathematics Volume 27, Berlin: de Gruyter.

[10] Frittelli M., Gianin E.R. (2002) *Putting order in risk measures*, Journal of Banking and Finance, Volume 26, Issue 7, Pages 1247-1561.

[11] Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. (2005) *Robust Statistics: the Approach Based on Influence Functions*, Chichester: Wiley.

[12] Huber P.J. (1981) *Robust Statistics*, Chichester: Wiley.

[13] Ibragimov R. and Walden J. (2001) *The limits of diversification when losses may be large*, Journal of Banking & Finance 31 (2007) 2551–2569.

[14] Kou S., Heyde C.C. and Peng X.H. (2007) *What Is a Good Risk Measure: Bridging the Gaps between Data, Coherent Risk Measures, and Insurance Risk Measures*, Financial Engineering Report 2007-09, Center for Financial Engineering, Columbia University.

[15] Kusuoka S. (2001) *On law invariant coherent risk measures*, Advances in Mathematical Economics, 3, p. 83-95

[16] Tasche D. (2002) *Expected Shortfall and Beyond.* Journal of Banking & Finance 26 1519-1533.

[17] Rockafellar R.T., Uryasev S. (2002) *Conditional Value-at-Risk for General Loss Distributions*, Journal of Banking and Finance, Volume 26, 7, 1443-1471.

[18] Van Zwet W.R. (1980) *A strong law for linear functions of order statistics.*, Ann. Proba. 8, 986-990.

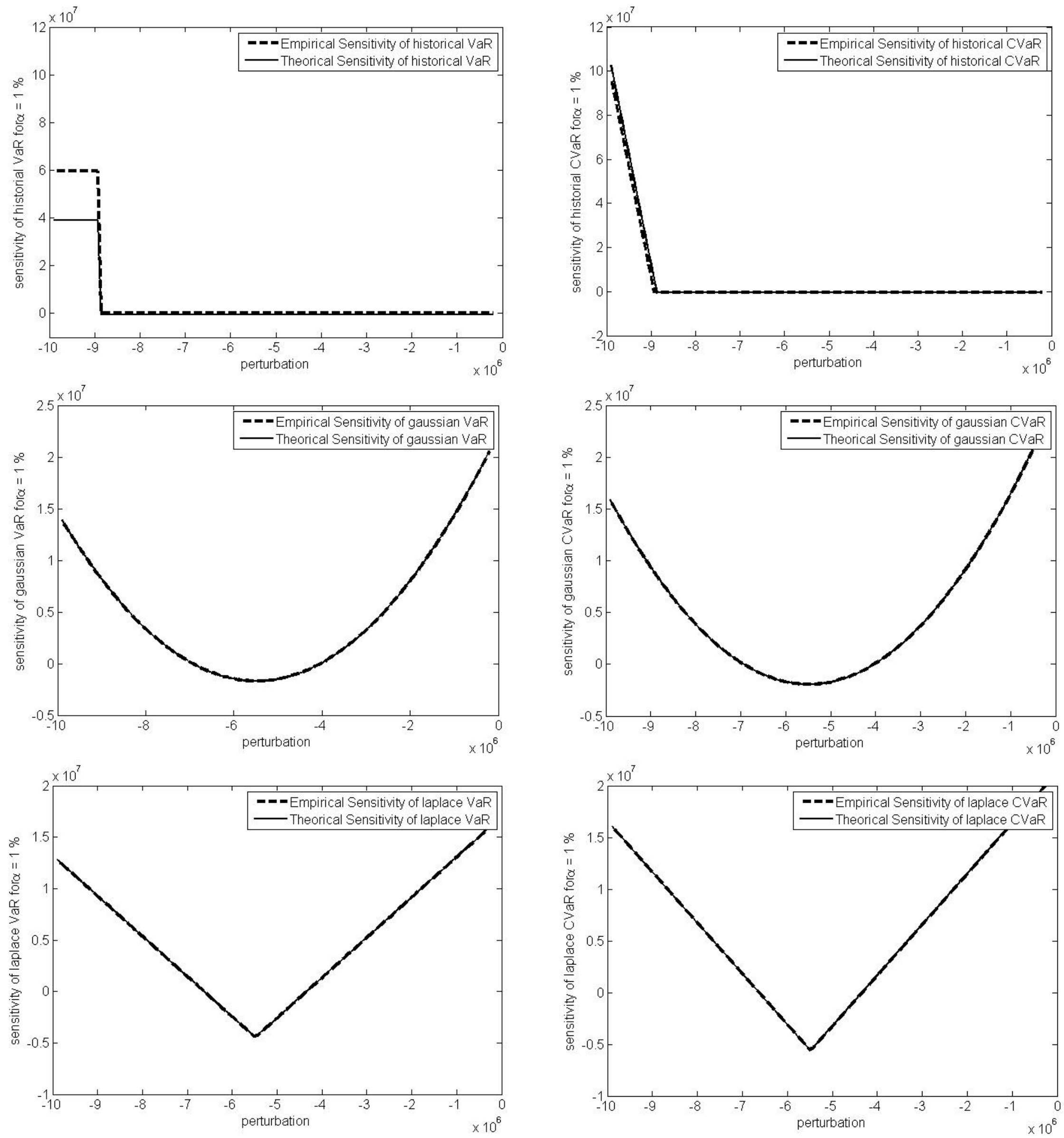



Figure 2: Sensitivity of risk estimators for $\alpha = 1\%$ at a 1 day horizon. Historical VaR (upper left), Historical CVaR (upper right), Gaussian VaR (left), Gaussian CVaR (right), Laplace VaR (lower left), Laplace CVaR (lower right).

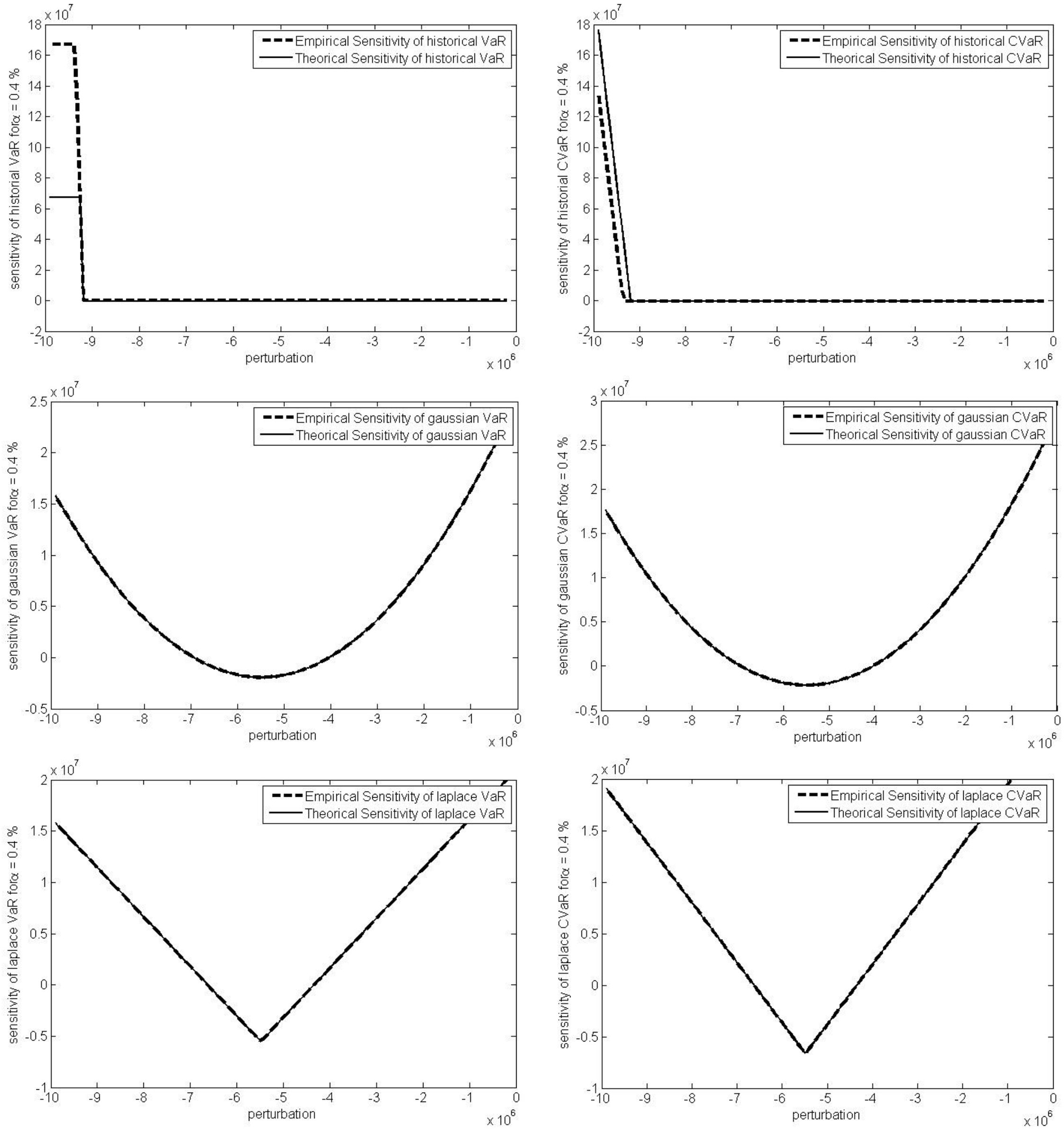



Figure 3: Sensitivity of risk estimators for $\alpha = 0.4\%$ at a 1 day horizon. Historical VaR (upper left), Historical CVaR (upper right), Gaussian VaR (left), Gaussian CVaR (right), Laplace VaR (lower left), Laplace CVaR (lower right).