Robust risk measurement from a Qualitative Point of View

Ennio Rapisarda

University of Catania

enniorapisarda@live.it

Abstract

Is the purpose of this paper to provide a contribution in order to understand what is a procedure of risk measurement (by axiomatic view) using a non-parametric approach and to determine which measures within this class are more robust than the others. To satisfy the scope we add some mathematical instruments and a numerical example to support theories.

Keywords: *Risk Measurement Procedure (RMP), Non parametric distributions, VaR, Expected Shortfall, Robust statistics.*

I. Introduction

LTo evaluate the riskiness of an investment portfolio, we must correlate the term "risk" to the *variability of the future value of a position*, due to market changes or more generally to uncertain events.

Here the risk can be defined as the *volatility* of unexpected outcomes, generally the value of assets or liabilities of interest. In our case we are focusing in *financial risks*, defined as those which relate to possible losses in financial markets, such as losses due to interest rate movements or default on financial obligations. Financial managers must necessarily understand risk, that is the chance to can consciously plan for the consequences of adverse outcomes and, by so doing, be better preparated for the inevitable uncertainty.

The increased volatility in exchange rates, interest rates and commodity prices has created the need for new financial instruments and analytical tools for risk management. Thus, *financial risk management* refers to the design and implementation of procedures for controlling financial risks, starting from the assumption that the typical operator on which we are referring, is a financial broker, and, as such, he is risk averse. Hence the main purpose in this outlook is to neutralize the uncertainty and the risk itself.

In 2010, Cont, Deguest and Scandolo provided a rigorous approach for examining how estimation issues can affect the computation of risk measures. Particularly, they focused on robustness and sensitivity analysis of risk measurement procedures, using tools from robust statistics. The main result was the fact that there is a conflict between coherence of a risk measure and the robustness of its commonly used estimators. We provide in this work a useful "tool-kit" to literature in order to better understanding how to construct a risk measure procedure, using a non-parametric approach, that may appear needful to supply a strong aspect to our results.

For this goal, in the first chapter we started to describe the probabilistic context and introducing the main kinds of risk measure both in analytical and practical terms.

In chapter two, we faced with statistical theory, that lead us to the notion of robustness and gives in output an instrument to quantify the risk by putting our attention on historical observations.

In chapter three we provide a real approach of robustness in order to verify what kind of measure satisfies our purposes, making use of a random experiment.

II. Monetary measures of risk and their coherent representation

2.1 A brief review of the probabilistic framework

We begin to state that a *random variable* (r.v.) is a measurable function from a probability space into the set of real numbers. An outcome (ω) for a r.v. is the mutually exclusive potential result that can occur. A *sample space* Ω is the set of all possible outcomes, and we can denote the fact that an outcome belongs to the sample space as $\omega_i \in \Omega$. An *event* is a subset of the sample space and can be represented as a collection of some outcomes. The collection of all events is usually denoted by $\mathcal{F}$, and it has to be considered together with Ω in order to denote the couple $(\Omega, \mathcal{F})$, just because the notion of probability is associated with an event.

We can identify, for random variables, two macro-categories:

- *Discrete*: limits the outcomes where the variable can only take on discrete values. The probabilistic treatment is comparatively easy. Once a probability is assigned to all different outcomes, the probability of an arbitrary event can be calculated by simply adding the single probabilities.
- *Continuous*: if can take on any possible value within the range of outcomes. The computation of probabilities is substantially different from the discrete case: the number of values in an interval is so large, that we cannot just add the probabilities of the single outcomes.

Now we define a *probability distribution function* P that assigns a probability $P(A)$ for every event A, that is, of realizing a value for the random value in any specified subset A on Ω. In the case of a continuous probability distribution, the most popular way to compute the probability is to provide the so-called *probability density function* as $f_X(x)$ for the r.v. X. This function is always non-negative and large values at some point x imply a relatively high probability of realizing a value in the neighborhood of x, whereas $f_X(x) = 0$ for all x in some interval (a, b) implies that the probability for observing a realization in (a, b) is zero.

Finally, the *cumulative distribution function* $F_X(x)$ is the mathematical function that provides the cumulative probability of a probability distribution, that is, the function that assigns to every real value x the probability of getting an outcome less than or equal to x. This function is always non-negative, non-decreasing and as it represents probabilities it takes only values between zero and one.

The mathematical connection between a probability density function f, a probability distribution P, and a cumulative distribution function F of some r.v. X is given by the following formula:

$$F_X(x) = P(X \leq t) = \int_{-\infty}^{t} f_X(x)dx.$$

Conversely, the density equals the first derivative of the distribution function,

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

A r.v. on the set of states of nature at a future date, can be interpreted as possible future value of positions of portfolio currently held. A first measurement of the risk of position will be whether its future value belongs or does not belongs to the subset of *acceptable risks* as decided by a supervisor (i.e. an investment manager). If a risk isn't contained in this subset, we can identify the measure of risk in terms of cash amount that is necessary to turn the position from "unacceptable" to acceptable. We consider only one period of uncertainty $(0, T)$ between two dates 0 and T, and we number the various rates of return by $i, 1 \leq i \leq I$ and, for each of them, one "reference instrument" is given, wich carries one unit of date 0 rate i into r_i units of date T rate i. For our purposes we can take into account Default free ZCB with maturity at date T.

An investor's initial portfolio consists of positions $A_i, 1 \leq i \leq I$, that provides $A_i(T)$

units of rate i at date T. We call *risk* the investor's *future net worth* $\sum_{1\leq i\leq I} e_i \cdot A_i(T)$.

We fix Ω as the set of possible outcomes of position that can be described by a list of the prices of all securities, and a mapping $X : \Omega \rightarrow \mathbb{R}$ where $X(\omega)$ is the discounted net worth of the position at T if the scenario $\omega \in \Omega$ is realized.

It will results needful to use *measures of dispersion* that allows us to describe a probability distribution function. In this case, the most commonly used measure is the *variance*, that measures the dispersion of the values that the r.v. can realize relative to the mean. Formally, is the average of the squared deviations from the mean, and is represented by squared units. Taking the root square of the variance, we obtain the *standard deviation*. In contrast to the variance, the mean absolute deviation takes the average of the absolute deviations from the mean.

2.2 Coherent risk and convex measures

We now introduce an instrument to quantify the risk of X as a number $\rho(X)$, where X belongs to a given class $\mathcal{X}$ of financial positions.

Definition 2.1 (Monetary measure of risk). *A mapping $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is called a monetary measure of risk if it satisfies the following conditions for all $X, Y \in \mathcal{X}$.*

- Monotonicity*: If $X \leq Y$, then $\rho(X) \geq \rho(Y)$. It means that the downside risk of a position is reduced if the payoff profile is increased.*
- Cash invariance*: If $m \in \mathbb{R}$, then $\rho(X + m) = \rho(X) - m$. Is also called* translation invariance*. $\rho(X)$ can be seen as the amount which should be added to the position X in order to make it acceptable: if the amount m is added to the position and invested in a risk-free manner, the capital requirement is reduced by the same amount. Moreover ensures that, for each X, $\rho(X + \rho(X) \cdot r) = 0$.*
- Normalization*: $\rho(0) = 0$*

Another axiom, necessary but not sufficient, to prevent concentration of risks to remain undetected is

- Relevance*: for all $X \in \mathcal{X}$ with $X \leq 0$ and $X \neq 0$, we have $\rho(X) > 0$.*

When positive, the number $\rho(X)$ assigned by the measure ρ to the risky stock (or position) X will be interpreted as the minimum extra cash the agent has to *add* to invest "prudently", that is in the reference instrument, to be allowed to proceed with his plans. If it is negative, the cash amount $-\rho(X)$ can be withdrawn from the position, or received as restitution as in the case of organized markets for financial future.

Now we need of another axiom, in order to analyse a special class of monetary measures:

Definition 2.2 (Convex measure of risk). *A monetary risk measure $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is called a convex measure of risk if it satisfies*

- Convexity*: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$, for $0 \leq \lambda \leq 1$.*

It means that, if ρ is convex and normalized, then

$$\begin{array}{lcll} \rho(\lambda X) & \leq & \lambda\rho(X) & \text{for} \quad 0 \leq \lambda \leq 1, \\ \rho(\lambda X) & \geq & \lambda\rho(X) & \text{for} \quad \lambda \geq 1, \end{array}$$

where λ and $(1-\lambda)$ represent the fraction of capital invested in respective strategies. Hence the diversification cannot increase the risk.

This kind of measures was identified because was found that in many situations the risk of a certain position might increase in a non-linear way with the size of the position.

Nevertheless, in our analysis we are going to focus in the following subset of convex measures of risk, identified by Artzner et al:

Definition 2.3 (Coherent measures of risk). *A monetary risk measure satisfying:*

- Sub-additivity*: for all X_1 and $X_2 \in \mathcal{X}$, $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$. This axiom represents a basic instrument through which we enforce the concept that "a merger does not create extra risk";*

- Positive homogeneity: *for all* $\lambda \geq 0$ *and* $X \in \mathcal{X}$, $\rho\lambda X) = \lambda\rho(X)$. *It means that when computing the future net worth of a position, we should consider the consequences of lack of liquidity (depending on its size);*

is called coherent.

Any coherent risk measure arises as the supremum of the expected negative of final net worth for some collection of "generalized scenarios" [1] or probability measures on states of the world.

Proposition 2.1. *Given the total return r on a reference instrument, a risk measure* ρ *is coherent if and only if there exists a family* $\mathcal{P}$ *of probability measures on the set of states of nature, s.t.*

$$\rho(X) = \sup\{\mathbf{E}_{\mathbb{P}}[-X/r] \mid \mathbb{P} \in \mathcal{P}\}.$$

The relevance axiom is satisfied by ρ if and only if the union of the supports of the probabilities in $\mathcal{P}$ is the whole set Ω of states of nature.

This measure, induces an *acceptance set* $\mathcal{A}_\rho$, that allow us to address the invariance of acceptability of a position with respect to a change of rates of return. How stated at the begin of this section, in order to identify how many cash amount we need to "jump" from an unacceptable position to a position suitable for our purposes, we have to consider a specific set $\mathcal{A}$. Hence in our processes regarding measurement of risk, we recognize:

Definition 2.4 (Risk measure associated to an acceptance set). *Given the total rate of return r on a reference instrument, the risk measure associated to the acceptance set* $\mathcal{A}$ *is the mapping from* $\mathcal{X}$ *to* $\mathbb{R}$ *denoted by* $\rho_{\mathcal{A},r}$ *and defined by*

$$\rho_{\mathcal{A},r}(X) = inf\{m \mid m \cdot r + X \in \mathcal{A}\}.$$

When we deal with market risk, we can interpret the state of the world as a list of the prices of all securities and all rates of return, then we assume that the set of all possible such lists is known. Furthermore we need of the assumption that markets at date T are liquid.

Definition 2.5 (Acceptance set associated to a risk measure). *The acceptance set associated to a risk measure* ρ *is the set denoted by* $\mathcal{A}_\rho$ *and defined by*

$$\mathcal{A}_\rho = \{X \in \mathcal{X} \mid \rho(X) \leq 0\}.$$

Before introducing the axioms related to this set, might be useful to locate some notation:

1. We remember that our r.v. X will represent a final net worth of a position, with negative part, $max(-X, 0)$ denoted by X^- and the supremum of X^- denoted by $\|X^-\|$. The r.v. identically equal to 1 is denoted by $\mathbf{1}_{\{\omega\}}$.
2. Since Ω is assumed to be finite, then $\mathcal{X}$ can be identified with $\mathcal{R}^n$, where $n = card(\Omega)$. The cone of non-negative elements in $\mathcal{X}$ shall be denoted by L_+, its negative by L_-.
3. We call $\mathcal{A}_{i,j}, j \in J_i$, a set of final net worths, expressed in rate i, which, in market i, are accepted by supervisor j.
4. We shall denote $\mathcal{A}_i$ the intersection $\bigcap_{j \in J_i} \mathcal{A}_{i,j}$ and use the generic notation $\mathcal{A}$ in the listing below.

Now we are able to state axioms for acceptance sets and to understand the reasons:

Axiom 2.1. *The acceptance set* $\mathcal{A}$ *contains* L_+.

(Non-negativity)

Axiom 2.2. *The acceptance set* $\mathcal{A}$ *does not intersect the set* L_{--} *where*

$$L_{--} = \{X \mid \omega \in \Omega, \quad X(\omega) < 0\}.$$

(No extra capital requested)

In this case, we shall consider a stronger condition for which the acceptance set $\mathcal{A}$ *satisfies* $\mathcal{A} \cap L_- = 0$.

[1] Under a given number of events, interpreted as down, up or no move for a future stock price, we can define "generalized scenarios" as the convex combination of the losses at the "extreme move" point and at the "no move at all" point.

The next axiom reflects risk aversion on the part of supervisor.

Axiom 2.3. *The acceptance set $\mathcal{A}$ is convex.*

A less netural requirement on the set of acceptable final net worth is:

Axiom 2.4. *The acceptance set $\mathcal{A}$ is a positively homogenous cone.*

Artzner et al [1] showed that, if a risk measure ρ is coherent, then the acceptance set is closed and satisfies the axioms 2.1, 2.2, 2.3. and 2.4. As a consequence, ρ satisfies the relevance axiom. From now on, X will denote the discounted random P&L.

2.3 Appropriate spaces of random payoffs

Let now introduce an approach that allows us to center our attention on the spaces where we have to employ studied theories. A complete, normed vector space is called a *Banach space*. The classical Banach spaces are the spaces of measurable functions $\mathcal{L}^p(\Omega, \mathcal{F}, P)$, $p \in [1; \infty]$, and the sequence spaces ℓ^p. The cases $p \in [1, \infty)$ and $p = \infty$ are quite different.

Definition 2.6 (**L^p spaces).** *For $p \in [1, \infty)$,*

- *$L^p = L^p(\Omega, \mathcal{F}, P)$ is the set of $\mathbb{R}$-valued r.v. with the norm*

$$\|X\|_p := (\int |X|^p dP)^{1/p} < \infty;$$

- *$L^\infty = L^\infty(\Omega, \mathcal{F}, P)$ is the set of r.v. with the norm*

$$\begin{aligned}\|X\|_\infty &:= ess \sup |X| \\ &:= \sup\{B : P(|X| < B) < 1\} < \infty.\end{aligned}$$

It is basically essential for us to mention the main results about these special spaces:

1. *Norm (or triangle) inequality*:$\| \cdot \|_p$ is a norm, and the associated metric, $d_p(X_1, X_2) = \|X_1 - X_2\|_p$, makes L^p a complete metric space. The norm inequality that we must show is $\|X_1 + X_2\|_p \leq \|X\|_p + \|Y\|_p$. This is because, with normed spaces, we use the metric $d(X, Y) = \|X - Y\|_p$.

   The proof of the requisite norm inequality is a fairly easy implication of *Jensen's Inequality*,

   Theorem 2.3.1 (Jensen). *If C is a convex subset of $\mathbb{R}^k, f : C \to \mathbb{R}$ is concave, and $X : \Omega \to C$ is an integrable r.v., then $Ef(X) \leq f(EX)$. If f is convex, then $Ef(X) \geq f(EX)$.*

   And now we can state the following theorem for norm inequality:

   Theorem 2.3.2 (Minkowski-Riesz). *For $X, Y \in L^p(\Omega, \mathcal{F}, P), p \in [1, \infty], \|X + Y\|_p \leq \|X\|_p + \|Y\|_p$.*

   An immediate implication of the norm equality is that *Banach Spaces* are locally convex. It means that when we draw diagrams of open balls, we inevitably make them convex.

   Lemma 2.1. *In any normed space, $B_\epsilon(x) = \{y : \|x - y\| < \epsilon\}$ is a convex set.*

   If U_x is an open subset of a normed space containing x, then there is an open convex set G_x with $G_x \subset U_x$, simply take G_x to be $B_\delta(x)$ for a sufficiently small δ. It is the failure of the property that makes us avoid the L^p spaces with $p \in (0, 1)$.

2. *Completeness*: Knowing that $L^p(\Omega, \mathcal{F}, P)$ is a normed space, we have to prove that is complete, which means that is a *Banach Space*, that is, a complete normed vector space.

   Theorem 2.3.3. *For $p \in [1, \infty], L^p(\Omega, \mathcal{F}, P)$ is a complete metric space.*

   The proof relies on the Borel-Cantelli theorem and Fatou's lemma.

2.4 "Law invariant" coherent risk measures

In the extension of [1] , Delbaen introduced two new instruments, that results mathematically convenient in order to handle risk measures on which we are interested:

Definition 2.7 (Submodular). *A mapping $\psi : L^\infty \to \mathbb{R}$ is called submodular if*

1. *For $X \leq 0$ we have that $\psi(X) \leq 0$.*
2. *If X and Y are bounded r.v. then $\psi(X+Y) \leq \psi(X) + \psi(Y)$.*
3. *For $\lambda > 0$ and $X \in L^\infty$ we have $\psi(\lambda X) = \lambda\psi(X)$.*

   The submodular function is called translation invariant if moreover
4. *For $X \in L^\infty$ and $a \in \mathbb{R}$ we have that $\psi(X+a) = \psi(X) + a$.*

Definition 2.8 (Supermodular). *A mapping $\theta : L^\infty \to \mathbb{R}$ is called supermodular if*

1. *For $X \leq 0$ we have that $\theta(X) \leq 0$.*
2. *If X and Y are bounded r.v. then $\theta(X+Y) \geq \theta(X) + \theta(Y)$.*
3. *For $\lambda > 0$ and $X \in L^\infty$ we have $\theta(\lambda X) = \lambda\theta(X)$.*

   The supermodular function is called translation invariant if moreover
4. *For $X \in L^\infty$ and $a \in \mathbb{R}$ we have that $\theta(X+a) = \theta(X) + a$.*

If ρ is a coherent risk measure and if we put $\theta(X) = -\rho(X)$, we obtain a *supermodular functional*. If μ is a purely finitely additive measure, the expression $\theta(X) = E_\mu[X]$ gives a translation invariant submodular function that cannot be described by a σ-additive probability measure. So we need extra conditions.

Definition 2.9. *The translation invariant supermodular mapping $\theta : L^\infty \to \mathbb{R}$ is said to satisfy the* Fatou property *if $\theta(X) \geq \limsup \theta(X_n)$, for any sequence, $(X_n)_{n\geq 1}$, of functions, uniformly bounded by 1 and converging to X in probability.*

The coherent risk measure ρ associated with the supermodular function θ satisfies the Fatou property if for the said sequences we have $\rho(X) \leq \liminf \rho(X_n)$. Using a characterisation of weak closed convex sets[2] in L^∞, we obtain:

Theorem 2.4.1. *For a translation invariant supermodular mapping θ, the following 4 properties are equivalent*

1. *There is an $L^1(\mathbb{P})$-closed, convex set of probability measures $\mathcal{Q}$, all of them being absolutely continuous with respect to $\mathbb{P}$ and such that for $X \in L^\infty$:*

$$\theta(X) = \inf_{\mathbb{Q}\in\mathcal{Q}} E_{\mathbb{Q}}[X];$$

2. *The convex cone $C = \{X \mid \theta(X) \geq 0\}$ is weak, i.e. $\sigma(L^\infty(\mathbb{P}), L^1(\mathbb{P}))$ closed;*
3. *θ satisfies the Fatou property;*
4. *If X_n is a uniformly bounded sequence that decreases to X a.s., then $\theta(X_n)$ tends to $\theta(X)$.*

In order to represent coherent risk measures through expected values, over a family of probabilities, we need of some "control measure". A set N is of *first category* if it is contained in the countable union of closed sets with empty interior. The class of Borel sets of first category, denoted by $\mathcal{N}$, forms a σ-ideal in $\mathcal{F}$. For a bounded function X defined on [0,1] and Borel measurable, we define $\rho(X)$ as the "essential" supremum of $-X$. More precisely we define

$$\rho(X) = \min\{m \mid \{-X > m\} \text{ is of first category}\}.$$

The associated submodular function is then defined as

$$\psi(X) = \min\{m \mid \{X > m\} \text{ is of first category}\}.$$

Thus, $\rho(X)$ defines a coherent risk measure and satisfies the Fatou property. Explained these prerequisites, we can introduce

[2] A set $K \subset L^p$ is *weakly closed* if it contains the limits of all weakly convergent sequences X_n in K.

Definition 2.10 (Law invariant risk measure). *A map $\rho : L^\infty \to \mathbb{R}$ is law invariant if $\rho(X) = \rho(Y)$ whenever $X, Y \in L^\infty$ have the same probability law.*

Let $\mathcal{D}$ be the set of probability distribution functions of bounded r.v., i.e.,$\mathcal{D}$ is the set of non-decreasing right-continuous functions F on $\mathbb{R}$ s.t. there are $z_0, z_1 \in \mathbb{R}$ for which $F(z) = 0, z < z_0$ and $F(z) = 1, z \geq z_1$. Let us define $Z : [0, 1) \times \mathcal{D} \to \mathbb{R}$ by

$$Z(x, F) \equiv q_X^+(t) = \inf\{x \mid F(x) > t\}, \qquad t \in [0, 1), \quad F \in \mathcal{D}.$$

To emphasize the dependence on F, we alternatively write $q_X^+(t) \equiv q_t^+(F) \equiv F^{\leftarrow}(t)$, where $F^{\leftarrow}(\cdot)$ denotes the generalized inverse of F. Then $q_\cdot^+(F) : [0, 1) \to \mathbb{R}$ is non decreasing and right-continuous. We denote by F_X the probability distribution function of a r.v. X. For each $\alpha \in (0, 1] \to \mathbb{R}$, let $\rho_\alpha : L^\infty$ be given by

$$\rho_\alpha(X) = \alpha^{-1} \int_{1-\alpha}^{1} q_t^+(F) dt, \qquad X \in L^\infty, \quad -X \sim F.$$

Also,we define $\rho_0 : L^\infty \to \mathbb{R}$ by

$$\rho_0(X) = ess\sup(-X), \qquad X \in L^\infty.$$

Then it is easy to see that $\rho_\cdot(X) : [0, 1] \to \mathbb{R}$ is a non-increasing continuous function for any $X \in L^\infty$. Assuming that $(\Omega, \mathcal{F}, P)$ is a standard probability space and P is non-atomic[3], we can state from [14] the following

Theorem 2.4.2. *Let $\rho : L^\infty \to \mathbb{R}$. Then the following conditions are equivalent.*

1. *There is a (compact convex) set $\mathcal{M}_0$ of probability measures on [0,1] s.t.*

$$\rho(X) = \sup\{\int_0^1 \rho_\alpha(X) m(d\alpha)\}; \qquad m \in \mathcal{M}_0, \quad X \in L^\infty \quad (1)$$

2. *ρ is a law invariant risk measure with the Fatou property.*

2.5 VaR

An example of measures of risk used in practice is the "Value at Risk" (VaR) measure, usually defined in terms of net wins or Profit/Losses and therefore ignores the differences between money at one date and money at a different date, which, for small time periods and a single rate of return, may be acceptable. VaR applies to any financial instrument and it is expressed in the same unit of measure, namely in "lost money". The comparison between,say, an equity portfolio and a forex portfolio is straight knowing their VaR's. Moreover, it includes an estimate of future events and allows one to convert in a single number the risk of a portfolio.

It uses quantiles [4] of the distribution of X under the given probability measure P.

Briefly, we remember that, for $\lambda \in (0, 1)$ (usually $\lambda \leq 10\%$), a $\lambda-$*quantile* of a r.v. X on $(\Omega, \mathcal{F}, P)$ is any real number q with the property

$$P(X \leq q) \geq \lambda \quad and \quad P(X < q) \leq \lambda,$$

and the set of all λ-quantiles of X is an interval $[q_X^-(\lambda), q_X^+(\lambda)]$, where

$$q_X^-(t) = \sup\{x \mid P(X < x) < t\} = \\ = \inf\{x \mid P(X \leq x) \geq t\} = x_{(\alpha)}$$

is the lower and

$$q_X^+(t) = \inf\{x \mid P(X \leq x) > t\} = \\ \sup\{x \mid P(X < x) \leq t\} = x^{(\alpha)}$$

is the upper quantile function of X (that we can see as a functional on a space of financial positions X). It is useful to notice that we can use the x-notation only if the dependence on X is evident, otherwise we must refer to the q-notion.

Definition 2.11 (Value at Risk). *Fix some level $\lambda \in (0, 1)$. For a financial position X, we define*

[3] A probability is non-atomic if one can partition the measure space arbitrarily finely.

[4] Which requires us to pay attention to discontinuities and *intervals* of quantile numbers.

its Value at Risk at level λ *as*

$$VaR_\lambda(X) := -q_X^+(\lambda) = q_{-X}^-(1-\lambda) = \\ = \inf\{m \mid P[X+m<0] \leq \lambda\}. \quad (2)$$

We can interpret $VaR_\lambda(X)$ as the smallest amount of capital which, if added to X and invested in the risk-free asset, holds the probability of a negative outcome below the λ level.

However, is relevant to notice some *limitations* of this type of measure:

- errors shall occur because the distributional assumptions may not correspond to the actual distribution of changes in the market factors;
- does not reflect the risk of low probability outcomes, even though these may entail large or even crippling losses;
- fails to satisfy the subadditivity property (hence does not encourage diversification, but rather we can face with an increasing of risk).

Furthermore, VaR is not a convex measure of risk. VaR is coherent only when it is based on the standard deviation of normal distributions[5]. The latter highlighted problem has been largely established by literature. Thereupon it could happen that a well-diversified portfolio requires more regulatory capital than a less diversified portfolio. Hence, managing risk by VaR may fail to stimulate diversification and it prevents to add up to the VaR of different risk sources. Anyway, we can introduce an applied method which we use for estimating the distribution of losses from the ZCB position (even if will be treated more detailed in the next section:

- *Non Parametric method.* Uses a number of historical observations of returns. The VaR is calculated as the empirical $(1-\alpha)$-percentile of these observations. An advantage of this method is that it does not make any particular assumption about distribution of returns. The obvious drawback is that it is not possible to extrapolate beyond the range of the data at hand.

2.6 Expected shortfall (ES)

Despite VaR does not satisfy some axioms that appears needful for investor's purposes, it remains a good tool for managing the risk. Is for this reason that financial engineering has identified some appliance to convert the nature of this kind of measure by its traditional form.

Differences may appear when the underlying loss distributions have discontinuities. In this case even the coherence property of ES can get lost unless one took care of the details in its definition. There is only one definition of ES which is robust in the sense of yielding a coherent risk measure regardless of the underlying distributions. Moreover, can be estimated effectively even in cases where the usual estimators for VaR fail.

Two properties play an important role:

- *insensitivity*: ES is continuous with respect to α, hence, regardless of the underlying distributions, one can be sure that the risk measured by ES_α will not change dramatically when there is a "switch" in the confidence level by - say - some base points.
- *monotonicity in α*: the smaller the level α, the greater is the risk.

In order to construct an "expansion" of risk measures, we can start by [2] from the following

Proposition 2.2. *Let ρ_i be risk measures for $i = 1,..,n$. Then, any convex combination $\rho = \sum_i \alpha_i \rho_i$, where $\alpha_i \in \mathbb{R}^+$ and $\sum_i \alpha_i = 1$, is a risk measure. Similarly, if ρ_α is a one-parameter family of risk measures $\alpha \in [a,b]$, then, for any measure $d\mu(\alpha)$ in $[a,b]$ with $\int_a^b d\mu(\alpha) = 1$, the* statistic *defined as*

$$\rho = \int_a^b d\mu(\alpha)\rho_\alpha$$

is a risk measure.

[5] for a normal distribution, VaR is proportional to the standard deviation.

Proof. One uses the requirement $\alpha_i > 0$ (or $d\mu(\alpha) > 0$) to check monotonicity and sub-additivity axioms, and the requirement $\sum_i \alpha_i = 1$ (or $\int_a^b d\mu(\alpha) = 1$) to check translation invariance axiom. □

Definition 2.12 (Expected Shortfall). *Let $F_X(x) = P(X \leq x)$ be the distribution function of the Profit & Loss X of a given portfolio π and let $F^{\leftarrow}(t)$ be the generalized inverse of F. The α-*Expected Shortfall *defined for $\alpha \in (0,1]$ as*

$$ES_{(\alpha)}(X) = -\frac{1}{\alpha}\int_0^\alpha F_X^{\leftarrow}(t)dt \qquad (3)$$

is a risk measure satisfying the coherence axioms.

For $\alpha = 0$ it is natural to extend (3) defining $ES_{(0)}(X)$ as "the very worst case scenario"

$$ES_{(0)}(X) = -ess \inf X.$$

This equation might be understood as the smallest coherent and law invariant risk measurement to dominate VaR. For continuous r.v. X, it coincides with $E[-X \mid -X \leq VaR_\alpha(X)]$.

If we denote by $\mathbf{1}_A = \mathbf{1}_A(a)$ the indicator function of the set A, i.e.

$$\mathbf{1}_A = \begin{cases} 0 & \text{if} \quad a \notin A \\ 1 & \text{if} \quad a \in A \end{cases}$$

and we consider $E[\max(0, -X)] < \infty$, we can state ES under this new form:

$$ES_\alpha(X) = -(1-\alpha)(E[X\mathbf{1}_{\{-X \geq q_\alpha(-X)\}}] + \\ + q_\alpha(-X)\{\alpha - P(-X < q_\alpha(X))\}. \qquad (4)$$

2.7 Spectral risk measures

Given some known risk measures, it is easy to generate a new risk measure[6], hence we take into account the previous construction of ES as the basis that allows us to put our attention on a new coherent measure. We introduce a measure $d\mu(\alpha)$ on $\alpha \in [0,1]$ under suitable integrability conditions. Proposition 2.2 ensures that the statistic

$$M_\mu(X) = \int_0^1 d\mu(\alpha)\alpha ES_\alpha(X) = \\ = -\int_0^1 d\mu(\alpha)\int_0^\alpha dp F_X^{\leftarrow}(p) \qquad (5)$$

is a risk measure as long as the normalization condition

$$\int_0^1 \alpha d\mu(\alpha) = 1 \qquad (6)$$

is satisfied. Now we are going to employ the Fubini-Tonelli theorem in order to interchange the integrals

$$M_\mu(X) = -\int_0^1 dp F_X^{\leftarrow}(p)\int_p^1 d\mu(\alpha) \equiv \\ \equiv -\int_0^1 dp F_X^{\leftarrow}(p)\phi(p) \equiv M_\phi(X) \qquad (7)$$

and so we can see that the parametrization in terms of any measure $d\mu(\alpha)$ can be traded with a parametrization in terms of a decreasing positive *risk spectrum* $\phi(p) = \int_p^1 d\mu(\alpha)$. The normalization condition translates into the following normalization condition for ϕ:

$$\int_0^1 \phi(p)dp = \int_0^1 dp\int_p^1 d\mu(\alpha) = \\ = \int_0^1 d\mu(\alpha)\int_0^\alpha dp = \int_0^1 d\mu(\alpha)\alpha = 1. \qquad (8)$$

Thus, for any measure $d\mu(\alpha)$ satisfying normalization (6), we have a different risk measure defined by (5) which can also be expressed by (7) with $\phi(p) = \int_p^1 d\mu(\alpha)$. Conversely, for any decreasing positive function $\phi(p) : [0,1] \to \mathbb{R}^+$ satisfying normalization (8), (7) provides a risk measure which can also be expressed by (5) with $d\mu(\alpha) = -d\phi(\alpha)$.

Some specific features, we have to talk about ϕ. First of all, it is an element of the normed space $\mathcal{L}^1([0,1])$, where every element is represented by a class of functions which differ at most on a subset of [0,1] of zero measure. The norm in this space is given by

$$\| \phi \| = \int_0^1 |\phi(p)|dp. \qquad (9)$$

[6] A convex combination of risk measures is coherent as well.

The properties of monotonicity and positivity of an element of $\mathcal{L}^1([0,1])$ cannot be defined pointwise as for functions. Hence, we adopt the following

Definition 2.13. *We will say that an element $\phi \in \mathcal{L}^1([a,b])$ is "positive" if $\forall I \subset [a,b]$*

$$\int_I \phi(p)dp \geq 0. \tag{10}$$

We will say that an element $\phi \in \mathcal{L}^1([a,b])$ is "decreasing" $\forall q \in (a,b)$ and $\forall \epsilon > 0$ s.t. $[q-\epsilon, q+\epsilon] \subset [a,b]$

$$\int_{q-\epsilon}^{q} \phi(p)dp \geq \int_{q}^{q+\epsilon} \phi(p)dp. \tag{11}$$

Now we can state the following

Definition 2.14 (Admissible risk spectrum). *An element $\phi \in \mathcal{L}^1([0,1])$ is said to be an "admissible" risk spectrum if*

1. *ϕ is positive;*
2. *ϕ is decreasing;*
3. *$\parallel \phi \parallel = 1$.*

From the above discussion we can therefore easily prove the following

Theorem 2.7.1. *Let $M_\phi(X)$ be defined by*

$$M_\phi(X) = -\int_0^1 F_X^{\leftarrow}(p)\phi(p)dp \tag{12}$$

with $\phi \in \mathcal{L}^1([0,1])$. If ϕ is an admissible risk spectrum then $M_\phi(X)$ is a risk measure.

Proof. For all admissible spectra $\phi \in \mathcal{L}^1([0,1])$ it is always possible to find a representative positive and decreasing function $\phi(p)$ which defines a measure μ on $[0,1]$ by $d\mu(\alpha) = -d\phi(\alpha)$. Then, the coherency of M_ϕ follows from (5) and (7) and proposition 2.1. □

III. Robust procedures

3.1 Statistical functional, plug-in and main properties: a small back-step

When we choose a risk measure among those that satisfy coherency, we can start the process that gives us in output our number, in order to quantify the entity of portfolio's risk. This process is divided in two steps, which the first is the *Estimation of the loss distribution F_X* with an empirical distribution obtained from a historical or simulated sample. We can formalize this step as a function from $\mathcal{X} = \cup_{n\geq 1}\mathbb{R}^n$ (the collection of all possible datasets) to $\mathcal{D}$; if $\mathbf{x} \in \mathcal{X}$ is a dataset, we denote $\hat{F}_{\mathbf{X}}$ the corresponding estimate of F_X.

Suppose that $X_1, \ldots, X_n$ are i.i.d. whit cdf F. We assume that F is completely unknown, subject only to some very general conditions such as continuity or existence of moments. The "parameter" $\theta = \theta(F)$ to be estimated (a real-valued function defined over this non-parametric class $\mathcal{F}$) is called *statistical functional* and it can be written as an expectation.

We have to consider the $\theta' s$ for which there exists an integer a and a function κ of a arguments s.t.

$$\theta = E[\kappa(X_1, \ldots, X_a)].$$

We can assume κ to be symmetric in its a arguments.

Now let us focus on the estimation of θ by means of n observations $X_1, ..., X_n$ from F, where we shall assume that $a \leq n$. Clearly, $\kappa(X_1, \ldots, X_n)$ is an *unbiased estimator* of θ and so is $\kappa(X_{i_1}, \ldots, X_{i_a})$ for any a-tuple

$$1 \leq i_1 < \cdots < i_a \leq n. \tag{13}$$

This shows that also

$$U = \frac{1}{\binom{n}{a}} \sum_{i_1,\ldots,i_a} \kappa(X_{i_1}, \ldots, X_{i_a}) \tag{14}$$

is an unbiased estimator for θ, where the sum extends over all a-tuples satisfying (13). This estimator results symmetric in the n variables $(X_1, \ldots, X_n)$. More exactly, it is the only symmetric estimator which is unbiased for all F for which $\theta(F)$ exists, and it can be shown to have smaller variance than any other such unbiased estimator.

For $a = 1$, $\theta = E\kappa(X_1)$ and

$$U = \frac{1}{n}\sum_{i=1}^{n} \kappa(X_i).$$

Since this is the average of n i.i.d. random variables, asymptotic normality follows from the classical central limit theorem provided $0 < Var(X_1) < \infty$.

We can see the sample mean, the sample median and the sample variance as consistent estimators of corresponding population quantity. Any such population quantity is a function of the distribution F of the X_i (assumed i.i.d.), and can therefore be written as $h(F)$, where h is a real-valued function defined over a collection $\mathcal{F}$ of distributions F. The mean $h(F) = E_F(X)$ is defined over the class $\mathcal{F}$ of all F with finite expectation.

To establish the connection between the sequence of sample statistics and the functional $h(F)$ that it estimates, we define the *sample cdf* $\hat{F}_n$ by

$$\hat{F}_n(x) = \frac{\text{Number of } X_i \leq x}{n}.$$

This is a cdf of a distribution that assign probability $1/n$ to each of the n sample values $X_1, X_2, \ldots, X_n$. The standard estimator of $h(F)$ based on n observations is equal to $h(\hat{F}_n)$, the *plug-in estimator* [7] of $h(F)$. Assuming that $h(F) = E_F(X)$, the expectation of a r.v. with cdf $\hat{F}_n$ is the sum of the probabilities $(1/n)$ multiplied by the values (X_i) taken on by a r.v. with cdf $\hat{F}_n$, i.e.,

$$h(\hat{F}_n) = \frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n = \bar{X}.$$

In the same way, when $h(F) = E_F[X - E(X)]^k$, it is seen that

$$h(\hat{F}_n) = \frac{1}{n}(X_1 - \bar{X})^k + \cdots + \frac{1}{n}(X_n - \bar{X})^k = \\ = \frac{1}{n}\sum(X_i - \bar{X})^k = M_k.$$

Now we introduce the integral

$$\int adF = \int a(x)dF(x) = E_F[a(X)]$$

so that in particular

$$\int adF = \begin{cases} \int a(x)f(x)dx \\ \text{when } F \text{ has a density } f \\ \sum a(x_i)P_F(X = x_i) \\ \text{when } F \text{ is discrete} \end{cases}$$

.

Thus, linking to our focus, we have as example

$$\int a(x)d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} ax_i.$$

3.2 Order statistics

On the basis of our kind of analysis related to the possible final net worths of portfolio under attention, it is natural to attribute a sort to the support of r.v., statistically viewed as the *ordered sample*

$$X_{n,n}, \leq \cdots \leq X_{1,n}.$$

Hence $X_{n,n} = \min(X_1, \ldots, X_n)$ and $X_{1,n} = M_n = \max(X_1, \ldots, X_n)$.
The r.v. $X_{k,n}$ is called the k^{th} *upper order statistic.* Let's identify the connection with the empirical cdf of a sample: for $x \in \mathbb{R}$ we introduce the *empirical cdf*

$$F_n(x) = \frac{1}{n}\text{card}\{i : 1 \leq i \leq n, X_i \leq x\} = \\ \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x\}}, \quad x \in \mathbb{R},$$

where I_A denotes the indicator function of the set A.

Now we can see that

$$X_{k,n} \leq x \qquad \text{if and only if} \qquad \sum_{i=1}^{n} I_{\{X_i > x\}} < k,$$

which implies that

$$P(X_{k,n} \leq x) = P\Big(F_n(x) > 1 - \frac{k}{n}\Big).$$

Upper order statistics estimate tails and quantiles, and also excess probabilities over certain thresholds.

[7] Plug-in estimators should generally have good asymptotic properties relative to the corresponding population measures because of the global asymptotic properties of the empirical cdf.

Proposition 3.1 (Cumulative distribution function of the k^{th} upper order statistics). *For $k = 1, \ldots, n$ let $F_{k,n}$ denote the cdf of $X_{k,n}$. Then:*

1. $F_{k,n}(x) = \sum_{r=0}^{k-1} \binom{n}{r} \bar{F}^r(x) F^{n-r}(x)$
2. *If F is continuous, then*

$$F_{k,n}(x) = \int_{-\infty}^{x} f_{k,n}(z) dF(z),$$

where

$$f_{k,n}(x) = \frac{n!}{(k-1)!(n-k)!} F_{n-k}(x) \bar{F}^{k-1}(x);$$

i.e. $f_{k,n}$ is a density of $F_{k,n}$ with respect to F.

We define the *joint density of k upper order statistics* through

Theorem 3.2.1. *If F is absolutely continuous with density f, then*

$$f_{X_{1,n},\ldots,X_{k,n}}(x_1, \ldots, x_k) = \\ = \frac{n!}{(n-k)!} F^{n-k}(x_k) \prod_{i=1}^{k} f(x_i), \\ x_k < \cdots < x_1.$$

3.3 Robustness principle and consistent estimator

Robustness signifies insensitivity to small deviations from the assumptions, or at least should impair the performance only slightly. More exactly, we refer to *distributional robustness*: the shape of the true underlying distribution deviates slightly from the assumed model.

The main goal is that somewhat larger deviations from the model should not cause a catastrophe.

Definition 3.1 (Non-parametric procedure). *A procedure is called non-parametric if it is supposed to be used for a wide, not-parametrized set of underlying distributions. Despite non-parametric, the sample mean is highly sensitive to outliers and therefore very non-robust.*

Definition 3.2 (Distribution-free test). *A test is called distribution-free if the probability of falsely rejecting the null hypothesis is the same for all possible underlying continuous distributions (optimal robustness of validity). This kind of test does not imply anything about the behavior of the power function.*

Many of the common test statistics and estimators depend on the sample $(x_1, \ldots, x_n)$ only through the empirical distribution function

$$F_n(x) = n^{-1} \sum 1_{\{x_i < x\}},$$

or, for more general sample spaces, through empirical measure

$$F_n = n^{-1} \sum \delta_{x_i}$$

where δ_x stands for the pointmass 1 at x, that is, we can write

$$T_n(x_1, \ldots, x_n) = T(F_n) \tag{15}$$

for some functional T defined on the space of empirical measures. Often T has a natural extension to a subspace of the space $\mathcal{M}$ of all probability measures on the sample space. For instance, if the limit in probability exists, put

$$T(F) = \lim_{n \to \infty} T(F_n), \tag{16}$$

where F is the true underlying common distribution of the observations. If a functional T satisfies (16), it is called *consistent* at F.

The natural robustness requirement for a statistic of the form (15) is that T should be continuous with respect to the *weak(-star)* topology, that is the weakest topology in the space $\mathcal{M}$ such that the map

$$F \to \int \eta dF \tag{17}$$

from $\mathcal{M}$ into $\mathbb{R}$ is continuous whenever η is bounded and continuous. If we take a linear statistic and make a small change in the sample, that is, make either small changes in all of the observations x_i or large changes in a few of them. If η is bounded and continuous, then this will result in a small change of $T(F_n) = \int \eta dF_n$. Conversely,

- if η is not bounded, then a single, strategically placed gross error can completely upset $T(F_n)$;

- if η is not continuous, and if F_n happens to put mass onto discontinuity points, then small changes in many of the x_i may produce a large change in $T(F_n)$.

According to Hampel (ref.) we can provide the following

Definition 3.3 (Robust sequences). *1. A sequence of estimators T_n is robust at a probability measure F if and only if*

$$\forall \quad \varepsilon > 0 \quad \exists \quad \delta > 0 \quad \forall G \quad \forall n:$$
$$\{\pi(F,G) < \delta \Rightarrow \pi(\mathcal{L}_F(T_n), \mathcal{L}_G(T_n)) < \varepsilon\}$$

2. A sequence T_n is robust in a neighborhood of F if and only if there is an $\iota > 0$ s.t. $\pi(F,G) < \iota$ implies T_n is robust at G. A sequence T_n is robust at a class $\mathcal{E} \subset \mathcal{M}$ if and only if it is robust at all $F \in \mathcal{E}$. A sequence T_n is robust (everywhere) iff it is robust at all $F \in \mathcal{M}$.

We provide a little change of notation: d in place of π, and we denote that d corresponds to the *Lévy Distance*

3.4 Historical risk estimators as L-functionals

Now, we consider a statistic that is a linear combination of order statistics, or more generally, of some function h of them:

$$T_n = \sum_{i=1}^{n} a_{ni} h(x_{(i)}). \tag{18}$$

In order to preserve the total mass, $\sum_i a_{ni} = M\{(0,1)\}$, and symmetry of the coefficients, if M is symmetric about $t = \frac{1}{2}$, we assume that the weights are generated by a signed measure M on $(0,1)$:

$$a_{ni} = \frac{1}{2} M\left\{ \left(\frac{i-1}{n}, \frac{i}{n} \right) \right\} + \frac{1}{2} M\left\{ \left[\frac{i-1}{n}, \frac{i}{n} \right] \right\}. \tag{19}$$

Then $T_n = T(F_n)$ derives from the functional

$$T(F) = \int h(F^{\leftarrow}(s)) M(ds). \tag{20}$$

In order to have exact equality, we have to regularize the integrand at its discontinuity point and to replace it by

$$\frac{1}{2} h(F_n^{\leftarrow}(s-0)) + \frac{1}{2} h(F_n^{\leftarrow}(s+0)), \tag{21}$$

but only asymptotic equivalence if we do not care. We remember that the inverse of any distribution function F is usually defined as

$$F^{\leftarrow}(s) = \inf\{x \mid F(x) \geq s\} \quad 0 < s < 1.$$

The second step of our procedure, consists in the *Application of the risk measure ρ*: to the estimated P&L distribution $\hat{F}_{\mathbf{x}}$, which yields an estimator $\hat{\rho}(\mathbf{x}) \triangleq \rho(\hat{F}_{\mathbf{x}})$ for $\rho(X)$. Finally, the combination of this step, with that seen in previous section gives us in output our

Definition 3.4 (Risk measurement procedure). *A* risk measurement procedure *(RMP) is a couple (M, ρ), 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.*

Thus, the outcome of this procedure is a *risk estimator* $\hat{\rho} : \mathcal{X} \to \mathbb{R}$ defined as

$$\mathbf{x} \mapsto \hat{\rho}(\mathbf{x}) \triangleq \rho(\hat{F}_{\mathbf{x}}),$$

that estimates $\rho(F_n(x))$ given the data(set) $\mathbf{x}$, according to the following diagram.

In this way, we obtain the *historical estimator $\hat{\rho}^h$* associated to a risk measure ρ by applying it to the sample cdf $\hat{F}_{\mathbf{x}} = F_{\mathbf{x}}^{emp}$:

$$\hat{\rho}^h(\mathbf{x}) = \rho(F_{\mathbf{x}}^{emp}).$$

Considering the following form of risk measure,

$$\rho_m(X) = -\int_0^1 q_u^-(X) m(du), \tag{22}$$

where m is a probability measure on (0,1), we have

$$\hat{\rho}_m^h(\mathbf{x}) = \rho_m(F_{\mathbf{x}}^{emp}) = -\sum_{i=1}^{n} w_{n,i} x_{(i)}, \quad \mathbf{x} \in \mathbb{R}^n,$$

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 $w_{n,n} = m\left(\frac{n-1}{n}, 1\right)$. From now on, we can easily derive the application of the historical estimators to the risk measures viewed in the 1^{st} chapter:

Definition 3.5 (Historical VaR). *Is given by*

$$\widehat{VaR}^h_\alpha(\mathbf{x}) = -x_{(\lfloor n\alpha\rfloor+1)}, \qquad (23)$$

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

Definition 3.6 (Historical Expected Shortfall). *Is given by*

$$\widehat{ES}^h_\alpha(\mathbf{x}) = -\frac{1}{n\alpha}\left(\sum_{i=1}^{\lfloor n\alpha\rfloor} x_{(i)} + \underbrace{x_{(\lfloor n\alpha\rfloor+1)}}_{=\widehat{VaR}^h_\alpha(\mathbf{x})}(n\alpha - \lfloor n\alpha\rfloor)\right) \qquad (24)$$

Definition 3.7 (Historical estimator of the spectral risk measure). *Is given by:*

$$\widehat{\rho}^h_\phi(\mathbf{x}) = -\sum_{i=1}^{n} w_{n,i}x(i),$$
$$\text{where} \quad w_{n,i} = \int_{(i-1)/n}^{i/n}\phi(u)du. \qquad (25)$$

In the sequel we will write $\rho_m(X) \equiv \rho_m(F)$ to emphasize the dependence on the P&L distributions F and G.

IV. Verifying qualitative robustness

4.1 $\mathcal{C}$-robustness of VaR and nonrobustness of spectral risks

We have to fix a set $\mathcal{C} \in \mathcal{D}$ that will represent a main reference: the set of conceivable loss distributions and $F \in \mathcal{C}$, assuming F not an isolated point of $\mathcal{C}$. $\mathcal{C}$-robustness at F takes over if a small perturbation of F (which stays in $\mathcal{C}$) results in a small change in the law of the estimators obtained from our IID sample with law F, i.e. the law of the estimator is continuous w.r.t. changes in F, *uniformly* in the size n of the data set. Clearly, the bigger is the set of perturbations $\mathcal{C}$, the harder is the for a risk estimator to be $\mathcal{C}$-robust.

Proposition 4.1. *Let ρ be a risk measure and $F \in \mathcal{C} \subseteq \mathcal{D}_\rho$. If $\widehat{\rho}^h$ is consistent with ρ at every $G \in \mathcal{C}$, then the following are equivalent:*

1. *the restriction of ρ to $\mathcal{C}$ is continuous (w.r.t. Lévy distance) at F;*
2. *$\widehat{\rho}^h$ is $\mathcal{C}$-robust at F.*

Proof. For any fixed $\varepsilon > 0$ and $G \in \mathcal{C}$, as $\widehat{\rho}^h$ is consistent with ρ at F and G, there exists $n^* \geq 1$ s.t.

$$d(\mathcal{L}_n(\widehat{\rho}^h, F), \delta_{\rho(F)}) + d(\mathcal{L}_n(\widehat{\rho}^h, F), \delta_{\rho(G)}) < \frac{2\varepsilon}{3},$$
$$\forall n \geq n*. \qquad (26)$$

- 1. $\Rightarrow$ 2. Assume that $\rho|_\mathcal{C}$ is continuous at F and fix $\varepsilon > 0$. Then there exists $\delta > 0$ s.t. if $d(F,G) < \delta$, then $d(\delta_{\rho(F)}, \delta_{\rho(G)}) = |\rho(F) - \rho(G)| < \varepsilon/3$. Thus $\mathcal{C}$-robustness readily follows from (26) and the triangular inequality
$$d(\mathcal{L}_n(\widehat{\rho}^h, F), \mathcal{L}_n(\widehat{\rho}^h, G)) \leq$$
$$\leq d(\mathcal{L}_n(\widehat{\rho}^h, F), \delta_{\rho(F)}) + d(\delta_{\rho(F)}, \delta_{\rho(G)}) \leq$$
$$\leq d(\mathcal{L}_n(\widehat{\rho}^h, G), \delta_{\rho(G)}).$$
- 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 $\bar{n} \geq 1$ s.t.
$$d(F,G) < \delta, \quad G \in \mathcal{C} \quad \Rightarrow$$
$$\leq d\big(\mathcal{L}_{\bar{n}}(\widehat{\rho}^h, F), \mathcal{L}_{\bar{n}}(\widehat{\rho}^h, G)\big) < \varepsilon/3.$$
As a consequence, from (26) and the triangular inequality
$$|\rho(F) - \rho(G)| = d(\delta_{\rho(F)}, \delta_{\rho(G)}) \leq$$
$$d\big(\delta_{\rho(F)}, \mathcal{L}_n(\widehat{\rho}^h, F)\big) + d\big(\mathcal{L}_n(\widehat{\rho}^h, F), \mathcal{L}_n(\widehat{\rho}^h, G)\big) +$$
$$+ d\big(\mathcal{L}_n(\widehat{\rho}^h, G), \delta_{\rho(G)}\big),$$
it follows that $\rho|_\mathcal{C}$ is continuous at F.

□

Now, we need a sufficient condition on our risk measure to ensure that the historical estimator is robust. Thus, following Proposition 4.1, we obtain the following

Corollary 4.1. *If ρ_m is continuous in $\mathcal{C}$ then ρ^h_m is $\mathcal{C}$-robust at any $F \in \mathcal{C}$.*

Proof. Fix $G \in \mathcal{C}$ and let $(X_n)_{n\geq 1}$ be an IID sequence distributed as G. □

Then, by Glivenko-Cantelli Theorem we have, for almost all ω

$$d(F^{emp}_{\mathbf{X}(\omega)}) \to \rho(G), \tag{27}$$

and therefore $\hat{rho}$ is consistent with ρ at G. The following important result of Cont et Al was adapted from Huber [12, Theorem 3.1]: for a measure m on $[0,1]$ let

$$\mathcal{A}_m \triangleq \{\alpha \in [0,1] : m(\{\alpha\}) > 0\}$$

the set of values where m puts a positive mass. We remark that $\mathcal{A}_m$ is empty for a continuous m as in the definition of spectral risk measures.

Theorem 4.1.1. *Let ρ_m be a risk measure of the form (22). If the support of m does not contain 0 nor 1 then ρ_m is continuous at any $F \in \mathcal{D}_\rho$ s.t. $q^+_\alpha(F) = q^-_\alpha(F)$ for any $\alpha \in \mathcal{A}_m$. Otherwise ρ_m is not continuous at any $F \in \mathcal{D}_\rho$.*

Substantially, this form of risk measure can be continuous at some F if and only if its computation does not involve any extreme quantile (close to 0 or 1). In this case, continuity is ensured provided F is continuous at all points where m has a point mass. Analysing the last corollary and the last theorem, we can put our attention on fundamental results applied to historical estimators:

- *Historical VaR⇒*

  being $\mathcal{A}_m = \{\alpha\}$, we obtain the following

  Proposition 4.2. *The historical risk estimator of VaR_α is $\mathcal{C}$-robust at any $F \in \mathcal{C}$, where*

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

  It means that if the quantile of the real loss distribution is uniquely determined, then the empirical quantile is a robust estimator.

- *Historical estimator of ES and spectral risk measures ⇒*

Proposition 4.3. *For any $F \in \mathcal{D}^p$, the historical estimator of ρ is $\mathcal{D}^p$-robust at F if and only if, for some $\varepsilon > 0$*

$$\phi(u) = 0 \quad \forall u \in (0,\varepsilon) \cup (1-\varepsilon, 1), \tag{28}$$

i.e. ϕ vanishes in a neighborhood of 0 and 1.

Proof. If (28) holds for some ε, then the support of m does not contain 0 nor 1. As $\mathcal{A}_m$ is empty, Theorem 3.1.1 yields continuity of ρ at any distribution in $\mathcal{D}_\rho$. Hence, we have $\mathcal{D}_\rho$-robustness of $\hat{\rho}$ at F thanks to Corollary 3.1.1.

In the opposite case, if (28) does not hold for any ε, then 0 or 1 (or both) are in the support of m and therefore ρ_ϕ is not continuous at any distribution in $\mathcal{D}_\rho$, in particular at F. Therefore, by Proposition 3.1.1 we can conclude that $\hat{\rho}$ is not $\mathcal{D}_\rho$-robust at F. □

An immediate and relevant consequence is the following

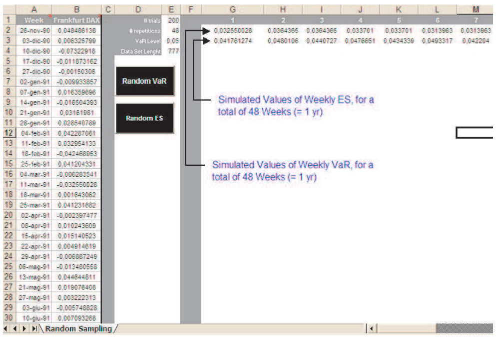
Corollary 4.2. *The historical risk estimator of any spectral risk measure ρ_ϕ defined on $\mathcal{D}^p$ is not $\mathcal{D}^p$-robust at any $F \in \mathcal{D}^p$. In particular, the historical risk estimator of ES_α is not $\mathcal{D}^1$-robust at any $F \in \mathcal{D}^1$.*

Proof. For a spectral risk measure, the density ϕ is decreasing and therefore it cannot vanish around 0, otherwise it would vanish on the entire interval $[0,1]$. □

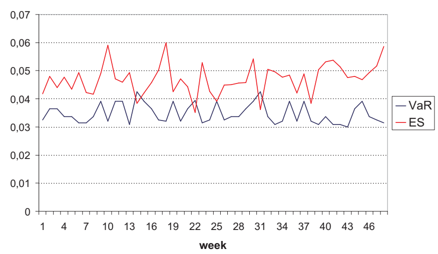
4.2 A numerical example

Recall from Section 2.4 that VaR is a robust risk functional even if it is not a coherent measure of risk. Instead, risk functionals based on the spectral representation (1.12) and (2.4) are nonrobust. Indeed, by Proposition 3.1.2 VaR is C-robust and by Corollary 3.1.2 ES is not. For the sake of completeness, we propose an empirical investigation to show the above results. More precisely, we use historical simulation to generate random samples from a dataset containing Frankfurt DAX quotes from 1990 to

2005. We coded a macro in VBA Microsoft Excel© to compute empirical estimators of VaR and ES. The following Figure illustrates the macro.



The following graph plots the values of the historical VaR and ES for a total of 48 replications per week.We can see how the overall path of historical VaR is more regular than that of ES, which is more volatile. Thus, we conclude that the historical estimator of ES is less robust, confirming the insight of Cont et al. [4].



V. Conclusion

VaR is a widespread tool in financial risk management. From a theoretical point of view, the recent literature on risk measures postulated that it is not coherent and discourage diversification to some extent. Alternative risk measures have been proposed and the most used from practitioners is ES. A central theme of modern risk management is the estimability of the P & L distribution function involved in the definition of the effective risk functionals. In this study we analyzed the robustness of VaR, spectral risk measures and ES as a special type. The $\mathcal{C}$-robustness proposed by Cont et al. [4] is the main result in risk measurement and it is based on a refined version of the classical Hampel's qualitative robustness. The surprising conclusion of this is that only the VaR is robust regardless its lack of coherence. We provide a numerical illustration confirming this result, encouraging future research in this direction.

V.References

[1] P. Artzner, F. Delbaen, J. Eber and D. Heat, *Coherent measures of risk*, Mathematical Finance, 9 (1999), pp. 203-228.

[2] C. Acerbi, *Spectral measures of risk: a coherent representation of subjective risk aversion*, Journal of Banking and Finance, 26 (2002), pp. 1505-1518.

[3] C. Acerbi and D. Tasche, *On the Coherence of Expected Shortfall*, Journal of Banking and Finance, 26 (2002), pp. 1487-1503.

[4] R. Cont, R. Deguest and G. Scandolo, *Robustness and Sensitivity Analysis of Risk Measurement Procedures*, Quantitative Finance, 6 (2010), pp. 593-606.

[5] D. Corbae, M.B. Stinchcombe and J. Zeman, *An Introduction to Mathematical Analysis for EconomicTheory and Econometrics*, Princeton University Press, 2009.

[6] A. DasGupta, *Asymptotic Theory of Statistic and Probability*, Springer, 2008.

[7] F. Delbaen, *Coherent Risk Measures on General Probability Spaces*, Preprint, 2000, E.T.H. Zurich.

[8] H. Föllmer and A. Schied, *Convex Measures of Risk and Trading Contraints*, Finance and Stochastics, 6 (2002), pp. 429-447.

[9] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, 2004.

[10] M. Frittelli and E. Rosazza Gianin, *Putting order in risk measures*, Journal of Banking and Finance, 26 (2002), pp. 1473-1486.

[11] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, *Robust Statistics - The Approach Based on Influence Functions*, Wiley, 1968.

[12] P. Huber, *Robust Statistics*, Wiley, 1981.

[13] S.G. Krantz, *Real Analysis and Foundation, 2nd edition*, Chapman & Hall/CRC, 2004.

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

[15] E.L. Lehmann, *Elements of Large-Sample Theory*, Springer, 2004

[16] D. Tasche, *Expected Shortfall and Beyond*, Statistical Data Analysis Based on the L1-Norm and Related Methods, (2002).