

# What Is a Good Risk Measure: Bridging the Gaps between Data, Coherent Risk Measures, and Insurance Risk Measures

C. C. Heyde, S. G. Kou, X. H. Peng  
Columbia University

June 26, 2006

## Abstract

Two main axiomatically based risk measures are the coherent risk measure, which assumes subadditivity for random variables, and the insurance risk measure, which assumes additivity for comonotonic random variables. We propose a new, data based, risk measure, called *natural risk statistic*, that is characterized by a new set of axioms. The new axioms only require subadditivity for comonotonic random variables, which is consistent with the prospect theory in psychology. Comparing to two previous measures, the natural risk statistic includes the tail conditional median which is more robust than the tail conditional expectation suggested by the coherent risk measure; and, unlike insurance risk measures, the natural risk statistics can also incorporate scenario analysis. The natural risk statistic includes the VaR as a special case and therefore shows that VaR, though simple, is not irrational.

**Keywords:** Risk measure, tail conditional expectation, tail conditional median, value at risk, quantile, robust statistics, L-statistics

## 1 Introduction

Broadly speaking a risk measure attempts to assign a single numerical value to the potential random financial loss. Obviously, it can be problematic in using one number to summarize the whole statistical distribution of the financial loss. Therefore, one shall avoid doing this if it is at all possible. However, in many cases there is no alternative choice. Examples of such cases include margin requirements in financial trading, insurance risk premiums, and government regulatory deposit requirements, such as Basel Accord [5, 6, 7] for banking regulation. Consequently, how to design a good risk measure becomes a problem of great practical importance, and is essential to financial and insurance industries as well as to the regulators.

There are two main approaches in the literature, the coherent risk measure suggested by Artzner et al. [3] and the insurance risk measure in Wang et al. [24]. In the coherent risk measure, one first chooses a set of scenarios (different probability measures) and then compute

the coherent risk measure as the maximal expectation of the loss under these scenarios. In the insurance risk measure, one fixes a distorted probability and then compute the insurance risk measure as the expectation with respect to one distorted probability. Both approaches are axiomatic, meaning that some axioms are postulated first, and all the risk measures satisfying the axioms are then identified.

Of course, once some axioms are set, there is room left to evaluate the axioms to see whether the axioms are reasonable for one's particular needs, and, if not, one should discuss possible alternative axioms. In particular, the subadditivity axiom in the coherent risk measure has been criticized by Daniélsson et al. [9] from a viewpoint of tail distribution, and by Dhaene et al. [11], Dhaene et al. [12] from regulatory and merger viewpoints.

In this paper we complement the previous approaches of coherent and insurance risk measures by postulating a different set of axioms. The resulting risk measures are fully characterized in the paper. More precisely, the contribution of the current paper is sixfold.

(1) We give more reasons on why a different set of axioms is needed: (a) In addition to reviewing the critiques of subadditivity in [9], [11], [12], we point out some new critiques from psychological and robustness viewpoints. See Section 3. (b) The main drawback of insurance risk measure is that it does not incorporate scenario analysis; i.e. unlike the coherent risk measure, the insurance risk measure chooses a (distorted) probability measure, and does not allow one to compare different distorted probability measures. See Section 2. (c) What is missed in both coherent and insurance risk measures is the consideration of data. Our approach is based on data (either observed or simulated or both) rather than on hypothetical distributions.

(2) A different set of axioms based on data and comonotonic subadditivity is postulated in Section 4, resulting in the definition of *natural risk statistic*. A complete characterization of the natural risk statistics is given in Theorem 1.

(3) An alternative characterization of the natural risk statistic based on statistical acceptance sets is given in Section 5.

(4) VaR or quantile is among the most widely used risk measures in practice (see, e.g. the Basil Accord). However the coherent risk measure rule out the use of VaR. In Section 6 we show that the natural risk statistic gives an axiomatic justification to the use of VaR.

(5) Unlike the insurance risk measures, we show in Section 6 that the natural risk statistic can incorporate scenario analysis by putting different weights on the sample order statistics.

(6) We point out in Section 7 that the natural risk statistic includes the *tail conditional median* as a special case, which leads to more robust measures of risk than that of the tail

conditional mean suggest by the coherent risk measure.

The mathematical difficulty of the current paper lies in the proof of Theorem 1. Unlike in the case of coherent risk measures, one cannot use the results in Huber [18] directly. This is because we only require comonotonic subadditivity, and the comonotonic sets are not open sets. Therefore, one has to be careful in applying the theorem of separating hyperplanes. Also we need to show that the weights are negative and add up to one.

## 2 Review of Two Risk Measures

### 2.1 Coherent Risk Measure

Let  $\Omega$  be the set of all possible states and  $\mathcal{X}$  be the set of all the random financial losses defined on  $\Omega$ . Then a risk measure is a mapping from  $\mathcal{X}$  to  $\mathbb{R}$ .

In Artzner et al. [3] a risk measure  $\rho$  is called a coherent risk measure, if it satisfies the following three axioms:

**Axiom A1.** Translation invariance and positive homogeneity:

$$\rho(aX + b) = a\rho(X) + b, \quad \forall a \geq 0, b \in \mathbb{R}.$$

**Axiom A2.** Monotonicity:  $\rho(X) \leq \rho(Y)$ , if  $X \leq Y$  almost surely.

**Axiom A3.** Subadditivity:  $\rho(X + Y) \leq \rho(X) + \rho(Y)$ , for any  $X, Y \in \mathcal{X}$ .

Axiom A1 states that the risk of a financial position is proportional to the size of the position and that a sure loss of amount  $b$  simply increases the risk by  $b$ . Axiom A2 is a minimum requirement for a reasonable risk measure. What is controversial lies in the subadditivity requirement in Axiom A3, which basically means that “a merger does not create extra risk” ([3], p. 209). We will discuss the controversies related to this axiom in Section 3.

If  $\Omega$  has *finite number of elements*, Artzner et al. [3] showed that a risk measure  $\rho$  is coherent if and only if there exists a family  $\mathcal{P}$  of probability measures on  $\Omega$ , such that

$$\rho(X) = \sup_{P \in \mathcal{P}} \{E^P[X]\}, \quad \forall X \in \mathcal{X},$$

where  $E^P[X]$  is the expectation of  $X$  under the probability measure  $P$ . Therefore, if  $\Omega$  has finite number of elements, then a coherent risk measure amounts to computing maximal expectation under different scenarios (different  $P$ 's), thus justifying scenarios analysis used in practise. Also if  $\Omega$  has finite number of elements, then Artzner et al. [3] also presented an equivalent approach of defining the coherent risk measure through the specification of acceptance sets, the sets of financial positions accepted by regulators or investors.

The coherent risk measure gives an axiomatic justification to the use of the “SPAN” system in calculating initial portfolio margin requirements at the Chicago Mercantile Exchange (see [8]). SPAN essentially defines a margin requirement as the maximal expected loss of the portfolio under 14 pre-specified scenarios.

However, from a practical viewpoint, a main drawback of the coherent risk measure is that it rules out the use of quantiles as risk measures, as it only includes expectations. One of the most widely used risk measures for regulators and internal risk management in banks is Value-at-Risk, or VaR for short, which is nothing but a quantile at some pre-defined probability level. More precisely, Given  $\alpha \in (0, 1)$ , the value-at-risk  $\text{VaR}_\alpha$  at level  $\alpha$  of the loss variable  $X$  is defined as the  $\alpha$ -quantile of  $X$ , i.e.,

$$\text{VaR}_\alpha(X) = \min\{x \mid P(X \leq x) \geq \alpha\}. \quad (1)$$

For example the banking regulation “Basil Accord” specifies a risk measure as VaR at 99 percentile.

Therefore, the very fact that coherent risk measures exclude VaR and quantiles post a serious inconsistency between the academic theory and industrial practise. The main reason of this inconsistency is due to the subadditivity in Axiom A3, which is a controversial axiom, as we will explain in Section 3. By relaxing this axiom and requiring subadditivity only for comonotonic random variables, we are able to find a new set of axioms in Section 4 which will include VaR and quantiles, thus eliminating this inconsistency.

## 2.2 Insurance Risk Measure

Insurance risk premiums can also be viewed risk measures, as they aim at using one numerical number to summarize future random financial losses. Wang et al. [24] proposed four axioms to characterize insurance risk premiums. More precisely, a risk measure  $\rho$  is said to be an insurance risk measure if it satisfies the following four axioms.

**Axiom C1.** Conditional state independence:  $\rho(X) = \rho(Y)$ , if  $X$  and  $Y$  have the same distribution. This means that the risk of a position is determined only by the loss distribution.

**Axiom C2.** Monotonicity:  $\rho(X) \leq \rho(Y)$ , if  $X \leq Y$  almost surely.

**Axiom C3.** Comonotonic additivity:

$$\rho(X + Y) = \rho(X) + \rho(Y), \text{ if } X \text{ and } Y \text{ are comonotonic,}$$

where random variables  $X$  and  $Y$  are comonotonic if and only if

$$(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0$$

holds almost surely for  $\omega_1$  and  $\omega_2$  in  $\Omega$ .

**Axiom C4.** Continuity:

$$\lim_{d \rightarrow 0} \rho((X - d)^+) = \rho(X^+), \quad \lim_{d \rightarrow \infty} \rho(\min(X, d)) = \rho(X), \quad \lim_{d \rightarrow -\infty} \rho(\max(X, d)) = \rho(X),$$

where  $(X - d)^+ = \max(X - d, 0)$ .

The notion of comonotonic random variables in axiom C3 is discussed by Schmeidler [25], Yaari [29] and Denneberg [10]. If two random variables  $X$  and  $Y$  are comonotonic, then  $X(\omega)$  and  $Y(\omega)$  always move in the same direction as the state  $\omega$  changes. Wang et al. [24] imposed axiom C3 based on the argument that the comonotonic random variables do not hedge against each other, leading to the additivity of the risks. However, this is only true if one focuses on one scenario. Indeed, if one have multiple scenarios, then the counterexample at the end of Section 6 will show that comonotonic additivity fail to hold.

Wang et al. [24] proved that if  $\mathcal{X}$  contains all the *Bernoulli*( $p$ ) random variables,  $0 \leq p \leq 1$ , then risk measure  $\rho$  satisfies axioms C1-C4 and  $\rho(1) = 1$  if and only if  $\rho$  has a Choquet integral representation with respect to a distorted probability

$$\rho(X) = \int X d(g \circ P) = \int_{-\infty}^0 (g(P(X > t)) - 1) dt + \int_0^{\infty} g(P(X > t)) dt, \quad (2)$$

where  $g(\cdot)$  is called the distortion function which is nondecreasing with  $g(0) = 0$  and  $g(1) = 1$ , and  $g \circ P(A) := g(P(A))$  is called the distorted probability. The detailed discussion of Choquet integration can be found in Denneberg [10].

It should be emphasized that VaR satisfies the axioms C1-C4 (see Corollary 4.6 in Denneberg [10] for a proof that VaR satisfying axiom C4) and henceforth is an insurance risk measure. But VaR is not a coherent risk measure, because it may not satisfy subadditivity (see [3]). In general, the insurance risk measure in (2) does not satisfy subadditivity, unless the distortion function  $g(\cdot)$  is concave (see Denneberg [10]).

A main drawback of insurance risk measure is that it does not incorporate scenario analysis. More precisely, unlike the coherent risk measure, the insurance risk measure chooses a fixed distortion function  $g$  and a fixed probability measure  $P$ , and does not allow one to compare different measures within a family  $\mathcal{P}$  of probability measures. This is inconsistent with industrial practices, as people generate different scenarios to get a suitable risk measure.

The main reason that the insurance risk measure rules out scenario analysis is that it requires comonotonic additivity. The counterexample at the end of Section 6 shows that even for comonotonic random variables, with different scenarios we may get strict subadditivity rather

than additivity. In our new approach in Section 4 we shall require comonotonic subadditivity instead of comonotonic additivity.

**Remark:** What is missing in both coherent and insurance risk measures is the consideration of data. Our approach in Section 4 is based on data (either observed or simulated or both) rather than on hypothetical distributions.

### 3 Is Subadditivity too Restrictive?

#### 3.1 Review of Existing Critiques

##### 3.1.1 Tail Subadditivity of VaR

VaR has been criticized because of its lack of subadditivity, except for elliptically distributed random vectors (see Embrechts et al. [14]). In fact, if  $Z = (Z_1, \dots, Z_n)$  is an  $n$ -dimensional random vector with an elliptical distribution, then for any two linear portfolio  $X = \sum_{i=1}^n a_i Z_i$  and  $Y = \sum_{i=1}^n b_i Z_i$ ,  $a_i, b_i \in \mathbb{R}$ , the subadditivity holds:  $\text{VaR}_\alpha(X + Y) \leq \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y)$ .

Although in the center of the distributions VaR may violate the subadditivity, in practice VaR is mostly used in tail regions. In an interesting paper Daniélsson et al. [9] showed that VaR is subadditive in the tail regions, provided that the tails are not too fat. More precisely, they proved that: (1) If  $X$  and  $Y$  are two asset returns having jointly regularly varying non-degenerate tails with tail index bigger than one, then there exists  $\alpha_0 \in (0, 1)$ , such that

$$\text{VaR}_\alpha(X + Y) \leq \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y), \forall \alpha \in (\alpha_0, 1).$$

(2) If the tail index of the  $X$  and  $Y$  are different, then a weaker form of tail subadditivity holds

$$\limsup_{\alpha \rightarrow 1} \frac{\text{VaR}_\alpha(X + Y)}{\text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y)} \leq 1.$$

Asset returns with tail index less than one have very fat tails; they are hard to find and easy to identify, and Daniélsson et al. [9] argued that they can be treated as special cases in financial modeling.

According to the Basel Accord, 99% VaR is the primary risk measurement tool for determining capital charges against market risk. Daniélsson et al. [9] also carried out simulations which show that  $\text{VaR}_\alpha$  is indeed subadditive when  $\alpha \in [95\%, 99\%]$  for most practical applications.

##### 3.1.2 Critique from the Viewpoint of Regulator

For a risky business like an insurance company with potential loss  $X$ , regulators impose a solvency capital requirement  $\rho(X)$  on the company to protect the insurance policyholder from

insolvency, where  $\rho$  is a risk measure. In other words, the company has to hold the capital  $\rho(X)$  in order to pay out the claims from the policyholders. The shortfall of a portfolio with loss  $X$  and solvency capital requirement  $\rho(X)$  is therefore  $\max(0, X - \rho(X)) = (X - \rho(X))^+$ , which is the part of the loss that cannot be covered by the insurer.

Dhaene et al. [11] suggested that from the point of view of the regulator, a merger of two portfolios  $X$  and  $Y$  should decrease the shortfall, i.e.

$$(X + Y - \rho(X + Y))^+ \leq (X - \rho(X))^+ + (Y - \rho(Y))^+$$

However, Dhaene et al. [11] showed that this condition implies *superadditivity* not *subadditivity*; and they also show that even the weaker requirement

$$E[(X + Y - \rho(X + Y))^+] \leq E[(X - \rho(X))^+] + E[(Y - \rho(Y))^+]$$

is violated by coherent risk measures, such as the tail conditional expectation.

### 3.1.3 Critique from the Merger Viewpoint

Subadditivity basically means that “a merger does not create extra risk” ([3], p. 209). However, Dhaene et al. [12] pointed out that many times merger may increase risk, particularly due to bankruptcy protections for each firms. For example, it is better to split a risky trading business into a separate sub-firm. This way, even if the loss from a sub-firm is enormous, the parent firm can simply let the sub-firm go bankrupt, thus confining the loss to that one sub-firm. Therefore, creating sub-firms may incur less losses compared to the situation that the parent firm does not split into sub-firms. In this sense merger may increase firm risk.

The famous example of the collapse of Britain’s Barings Bank (which has a long operating history and even helped finance the Louisiana Purchase by the United States in 1802) in February 1995 due to the failure of a single trader, Nick Leeson, in Singapore clearly indicates that merger may increase risk. Had Barings Bank set up a separate firm for its Singapore unit, the bankruptcy in that unit would not have sink the entire bank.

## 3.2 Two New Critiques

### 3.2.1 Critique from the Psychological Theory of Uncertainty and Risk

Risk measures have a close connection with the psychological theory of people’s preference of uncertainties and risk. By the late 1970’s, psychologists and economists have discovered many preference anomalies in which rational people may systematically violate the axioms of

the expected utility theory. In particular, Kahneman and Tversky [19] proposed a model of choice under uncertainty called the “prospect theory” which departed from the expected utility model, leading to a Nobel prize in Economics. The theory has been further developed by many people, including the contributions from Kahneman and Tversky [27], Tversky and Wakker [28], Quiggin [21], Schmeidler [25], [26] and Yaari [29] on using comonotonic random variables in prospect theory. Their models are also referred to as “anticipated utility”, “rank-dependent models” and “Choquet expected utility”.

The prospect theory postulates that (a) it is better to impose preference on comonotonic random variables rather than on arbitrary random variables; and (b) people evaluate uncertain prospects using “decision weights” that may be viewed as distorted probabilities of outcomes. The theory can explain a variety of preference anomalies including the Allais and Ellsberg paradoxes.

Here is a simple example, by modifying a similar one in Schmeidler [26], that illustrates that risk associated with non-comonotonic random variables may violate subadditivity, because of Ellsberg’s paradox.

Suppose there are two urns I and II. Urn I contains 50 black balls and 50 white balls, while Urn II has unknown number of red and green balls. Let IB be the event that a ball drawn from Urn I is black, and let IW, IIR, IIG be defined similarly. The participants in the experiment are asked about their judgements of risk preference of losing 10,000 if one of the four events IB, IW, IIR, IIG happens. Then almost all participants seem to agree that the risk of IB and IW should be the same, and so does IIR and IIG. Furthermore, if we assign 10,000 as a risk measure for the sure loss of 10,000 with probability one, then people will have no problem accepting that the risk measures for IB and IW are all equal to 5,000. However, the paradox is that for a significant number of rational participants (for example people who know calculus, etc.), due to uncertainty in Urn II and people’s attitude towards large loss (p. 298 in Tversky and Kahneman [27]), they will prefer the loss of 10,000 associated with the event IIR over that with the event IB, and the event IIG over IW.

In other words, these rational participants view the risk of loss associated with IIR to be smaller than that with IB, and similarly for IIG and IW. Mathematically we have

$$\rho(IB) > \rho(IIR), \quad \rho(IW) > \rho(IIG), \quad \rho(IB) = 5000, \quad \rho(IW) = 5000.$$

But  $IIR \cup IIG$  corresponds a sure loss of 10,000, which leads to that

$$\rho(IIR \cup IIG) = \rho(10,000) = 10,000 = \rho(IB) + \rho(IW) > \rho(IIR) + \rho(IIG), \quad (3)$$

violating the subadditivity. Clearly the random losses associated with *IIR* and *IIG* are not comonotonic. Therefore, this example shows that non-comonotonic random variables may not have subadditivity.

Schmeidler [26] attributes this phenomena, such as the inequality in (3), to the difference between randomness and uncertainty, and further postulates that even for rational decision makers their subjective probabilities may not add up to one, due to uncertainty. Schmeidler [26] also indicated that risk preference for comonotonic random variables are easier to justified than the risk preference for arbitrary random variables

Following the prospect theory, we think it may be appropriate to relax the subadditivity to comonotonic subadditivity. In other words, we impose  $\rho(X + Y) \leq \rho(X) + \rho(Y)$  only for comonotonic random variables  $X$  and  $Y$ .

The insurance risk measure imposes comonotonic additivity in axiom C3 based on the argument that comonotonic losses have no hedge effect against each other. However, this intuition only holds when one focuses only on one scenario or one distorted probability. The counterexample at the end of Section 6 shows that if one incorporates different scenarios, then additivity may not hold even for comonotonic random variables. Hence, the comonotonic additivity condition in Axiom C3 may be too restrictive and its relaxation to comonotonic subadditivity may be a better choice.

### 3.2.2 Critique from the Robustness Viewpoint

When a regulator imposes a risk measure based on either probabilistic models of the loss or historical data of the loss, the risk measure must be unambiguous, stable, and must give consistent answers. Otherwise, different firms using different models may report very different risk measures to the regulator, and the risk measures may also fluctuate dramatically when firms update their historical database. In short, from a regulator viewpoint, the risk measure should demonstrate robustness with respect to underlying models or updating of historical data, in order to enforce the regulation and to maintain the stability of the regulation.

The robustness of the coherent risk measure is questionable:

(1) The coherent risk measure suggests to use tail expectations to compute risk measures, and tail expectations may be sensitive to model assumptions of heaviness of tail distributions, which is a controversial subject. For example, although it is accepted that real returns data has tails heavier than those of normal distribution, one school of thought believes tails to be exponential type (e.g. [13] [4]) and another believes power-type tails (e.g. [22]).

Heyde and Kou [17] shows that it is very difficult to distinguish between exponential-type and power-type tails with 5,000 observations (about 20 years of daily observations). This is mainly because the quantiles of exponential-type distributions and power-type distributions may overlap. For example, surprisingly, an exponential distribution has larger 99 percentile than the corresponding  $t$ -distribution with degree of freedom 5. Therefore, with ordinary sample data (e.g. 20 years of daily data), one cannot easily identify exact tail behavior from data. Therefore, the tail behavior may be a subjective issue depending on people's modeling preference.

We show in Section 7 that the tail conditional expectation, a widely used coherent risk measures, is sensitive to the assumption on the tail behavior of the loss distribution.

(2) Some risk measures may be coherent, satisfying subadditivity, but not robust at all. A simple example is the sample maxima. More precisely, given a set of observations  $\tilde{x} = (x_1, \dots, x_n)$  from a loss random variable  $X$ , let  $(x_{(1)}, \dots, x_{(n)})$  denote the order statistics of the data  $\tilde{x}$  with  $x_{(n)}$  being the largest. Then  $x_{(n)}$  is a coherent risk measure, as it satisfies subadditivity. However, the maximum loss  $x_{(n)}$  is not robust at all, and is quite sensitive to both outliers and model assumptions.

More generally, let  $\tilde{w}^0 = (w_1^0, \dots, w_n^0) \in \mathbb{R}^n$  be a weight with  $0 \leq w_1^0 \leq w_2^0 \leq \dots \leq w_n^0$  and  $\sum_{i=1}^n w_i^0 = 1$ . Then the risk measure  $\hat{\rho}(\tilde{x}) = \sum_{i=1}^n w_i^0 x_{(i)}$  is an empirically coherent risk measure satisfying subadditivity, as will be shown in Section 6. However, since this risk measure puts larger weights on larger observations, it is obviously not robust with respect to outliers in the data and model assumptions. We will discuss the issue of robustness in detail in Section 7.

## 4 Natural Risk Statistic and its Representation

### 4.1 The Axioms and the Representation

In this section we shall propose a new measure of risk based on data. Suppose we have a collection of data observation  $\tilde{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  on the random variable  $X$ . The collection  $\tilde{x}$  may be a set of empirical data observations from  $X$ , or a set of simulated data observation regarding (subjective) possible outcomes of  $X$  from a given model, or a combination of the two. Our risk measure, call natural risk statistic, is based on the data  $\tilde{x}$ . More precisely, a risk statistic  $\hat{\rho}$  is a mapping from the data in  $\mathbb{R}^n$  to a numerical value in  $\mathbb{R}$ .

**Remark:** In the setting of the coherent risk measure,  $\Omega$  has to be finite and consequently the random variable  $X$  has to be discrete. In our setting of risk statistic,  $X$  can be any random variable, discrete or continuous. What we need is a set of data observation (could be empirical

or subjective or both)  $\tilde{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  from  $X$ .

Next we shall introduce a set of axioms for  $\hat{\rho}$ .

**Axiom D1.** Positive homogeneity and translation invariance:

$$\hat{\rho}(a\tilde{x} + b\mathbf{1}) = a\hat{\rho}(\tilde{x}) + b, \quad \forall \tilde{x} \in \mathbb{R}^n, \quad a \geq 0, \quad b \in \mathbb{R},$$

where  $\mathbf{1} = (1, 1, \dots, 1)^T \in \mathbb{R}^n$ .

**Axiom D2.** Monotonicity:

$$\hat{\rho}(\tilde{x}) \leq \hat{\rho}(\tilde{y}), \quad \text{if } \tilde{x} \leq \tilde{y},$$

where  $\tilde{x} \leq \tilde{y}$  if and only if  $x_i \leq y_i, i = 1, \dots, n$ .

The above two axioms have been proposed for the coherent risk measure. Here we simply adapted them to the case of risk statistic. Note that Axiom D1 yields

$$\hat{\rho}(0 \cdot \mathbf{1}) = 0, \quad \hat{\rho}(b\mathbf{1}) = b, \quad b \in \mathbb{R},$$

Also Axioms D1 and D2 imply the continuity of  $\hat{\rho}$ . Indeed, suppose  $\hat{\rho}$  satisfies Axioms D1 and D2. Then for any  $\tilde{x} \in \mathbb{R}^n$ ,  $\varepsilon > 0$ , and  $\tilde{y}$  satisfying  $|y_i - x_i| < \varepsilon, i = 1, \dots, n$ , we have  $\tilde{x} - \varepsilon\mathbf{1} < \tilde{y} < \tilde{x} + \varepsilon\mathbf{1}$ . By the monotonicity in Axiom D2, we have  $\hat{\rho}(\tilde{x} - \varepsilon\mathbf{1}) \leq \hat{\rho}(\tilde{y}) \leq \hat{\rho}(\tilde{x} + \varepsilon\mathbf{1})$ . Applying Axiom D1, the inequality further becomes  $\hat{\rho}(\tilde{x}) - \varepsilon \leq \hat{\rho}(\tilde{y}) \leq \hat{\rho}(\tilde{x}) + \varepsilon$ , which establishes the continuity of  $\hat{\rho}$ .

**Axiom D3.** Comonotonic subadditivity:

$$\hat{\rho}(\tilde{x} + \tilde{y}) \leq \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y}), \quad \text{if } \tilde{x} \text{ and } \tilde{y} \text{ are comonotonic,}$$

where  $\tilde{x}$  and  $\tilde{y}$  are comonotonic if and only if  $(x_i - x_j)(y_i - y_j) \geq 0$ , for any  $i \neq j$ .

In Axiom D.3 we relax the subadditivity requirement in the coherent risk measure so that the axiom is only enforced for comonotonic data. This also relaxes the comonotonic additivity requirement in the insurance risk measure. Comonotonic subadditivity is consistent with the prospect theory of risk in psychology, as we only specify our preference among comonotonic random variables.

**Axiom D4.** Permutation invariance:

$$\hat{\rho}((x_1, \dots, x_n)) = \hat{\rho}((x_{i_1}, \dots, x_{i_n})), \quad \text{for any permutation } (i_1, \dots, i_n).$$

This axiom is postulated because we focus on risk measures of a single random variable  $X$  with data observation  $\tilde{x}$ . In other words, just like the coherent risk measure and insurance risk measure, here we discuss static risk measures rather than dynamic risk measures.

**Definition 1.** A risk statistic  $\hat{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}$  is called a natural risk statistic if it satisfies Axioms D1-D4.

The following representation theorem for natural risk statistic is a main result of the current paper.

**Theorem 1.** Let  $x_{(1)}, \dots, x_{(n)}$  be the order statistics of the observation  $\tilde{x}$  with  $x_{(n)}$  being the largest. Then  $\hat{\rho}$  is a natural risk statistic if and only if there exists a set of weights  $\mathcal{W} = \{\tilde{w} = (w_1, \dots, w_n)\} \subset \mathbb{R}^n$  with each  $\tilde{w} \in \mathcal{W}$  satisfying  $\sum_{i=1}^n w_i = 1$  and  $w_i \geq 0, \forall 1 \leq i \leq n$ , such that

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_{(i)} \right\}, \quad \forall \tilde{x} \in \mathbb{R}^n. \quad (4)$$

The main difficulty in proving Theorem 1 is the “only if” part. Axiom D3 implies that the functional  $\hat{\rho}$  satisfies subadditivity on comonotonic sets of  $\mathbb{R}^n$ , for example, on the set  $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \leq y_2 \leq \dots \leq y_n\}$ . However, unlike in the case of the coherent risk measure, the existence of the set of weights  $\mathcal{W}$  such that (4) holds does not follow easily from the proof in Huber [18]. The main difference here is that the comonotonic set  $\mathcal{B}$  is not an open set in  $\mathbb{R}^n$ . The boundary points may not have nice properties as the interior points do. We have to treat boundary points with more efforts. In particular, one should be cautious when using the results of separating hyperplanes. Furthermore, we have to spend some effort in Lemma 2 to show that

$$w_i \geq 0, \forall 1 \leq i \leq n.$$

## 4.2 Proof of Theorem 1

The proof relies on the following two Lemmas, which depend heavily on the properties of interior points. Therefore, we can only show that they are true for the interior points of  $\mathcal{B}$ . The results for boundary points will be obtained by approximating the boundary points by the interior points, and by employing continuity and uniform convergency.

**Lemma 1.** Let  $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \leq y_2 \leq \dots \leq y_n\}$ , and denote  $\mathcal{B}^\circ$  to be the interior of  $\mathcal{B}$ . For any fixed  $\tilde{z} = (z_1, \dots, z_n) \in \mathcal{B}^\circ$  and any  $\hat{\rho}$  satisfying Axioms D1-D4 and  $\hat{\rho}(\tilde{z}) = 1$  there exists a weight  $\tilde{w} = (w_1, \dots, w_n)$  such that the linear functional  $\lambda(\tilde{x}) := \sum_{i=1}^n w_i x_i$ , satisfies

$$\lambda(\tilde{z}) = 1, \quad (5)$$

$$\lambda(\tilde{x}) < 1 \text{ for all } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x}) < 1. \quad (6)$$

*Proof.* Let  $U = \{\tilde{x} \mid \hat{\rho}(\tilde{x}) < 1\} \cap \mathcal{B}$ . Since  $\tilde{x}, \tilde{y} \in \mathcal{B}$ , we know that  $\tilde{x}$  and  $\tilde{y}$  are comonotonic, Axioms D1 and D3 imply that  $U$  is convex, and, therefore, the closure  $\bar{U}$  of  $U$  is also convex.

For any  $\varepsilon > 0$ , since  $\hat{\rho}(\tilde{z} - \varepsilon \mathbf{1}) = \hat{\rho}(\tilde{z}) - \varepsilon = 1 - \varepsilon < 1$ , it follows that  $\tilde{z} - \varepsilon \mathbf{1} \in U$ . Since  $\tilde{z} - \varepsilon \mathbf{1}$  tends to  $\tilde{z}$  as  $\varepsilon \downarrow 0$ , we know that  $\tilde{z}$  is a boundary point of  $U$  because  $\hat{\rho}(\tilde{z}) = 1$ . Therefore, there exists a supporting hyperplane for  $\bar{U}$  at  $\tilde{z}$ , i.e., there exists a nonzero vector  $\tilde{w} = (w_1, \dots, w_n) \in \mathbb{R}^n$  such that  $\lambda(\tilde{x}) := \sum_{i=1}^n w_i x_i$  satisfies  $\lambda(\tilde{x}) \leq \lambda(\tilde{z})$  for all  $\tilde{x} \in \bar{U}$ . In particular, we have

$$\lambda(\tilde{x}) \leq \lambda(\tilde{z}), \forall \tilde{x} \in U. \quad (7)$$

We shall show that the strict inequality holds in (7). Suppose, by contradiction, that there exists  $\tilde{x}^0 \in U$  such that  $\lambda(\tilde{x}^0) = \lambda(\tilde{z})$ . For any  $\alpha \in (0, 1)$ , let  $\tilde{x}^\alpha = \alpha \tilde{z} + (1 - \alpha) \tilde{x}^0$ . Then we have

$$\lambda(\tilde{x}^\alpha) = \alpha \lambda(\tilde{z}) + (1 - \alpha) \lambda(\tilde{x}^0) = \lambda(\tilde{z}) \quad (8)$$

In addition, since  $\tilde{z}$  and  $\tilde{x}^0$  are comonotonic (as they all belong to  $\mathcal{B}$ ) we have

$$\hat{\rho}(\tilde{x}^\alpha) \leq \alpha \hat{\rho}(\tilde{z}) + (1 - \alpha) \hat{\rho}(\tilde{x}^0) < \alpha + (1 - \alpha) = 1, \quad \forall \alpha \in (0, 1). \quad (9)$$

Since  $\tilde{z} \in \mathcal{B}^\circ$ , it follows that there exists a small enough  $\alpha_0 \in (0, 1)$  such that  $\tilde{x}^{\alpha_0}$  is also an interior point of  $\mathcal{B}$ . Hence, for all small enough  $\varepsilon > 0$ ,

$$\tilde{x}^{\alpha_0} + \varepsilon \tilde{w} \in \mathcal{B}. \quad (10)$$

With  $w_{\max} = \max(w_1, w_2, \dots, w_n)$ , we have  $\tilde{x}^{\alpha_0} + \varepsilon \tilde{w} \leq \tilde{x}^{\alpha_0} + \varepsilon w_{\max} \mathbf{1}$ . Thus, the monotonicity in Axiom D2 and translation invariance in Axiom D1 yield

$$\hat{\rho}(\tilde{x}^{\alpha_0} + \varepsilon \tilde{w}) \leq \hat{\rho}(\tilde{x}^{\alpha_0} + \varepsilon w_{\max} \mathbf{1}) = \hat{\rho}(\tilde{x}^{\alpha_0}) + \varepsilon w_{\max}. \quad (11)$$

Since  $\hat{\rho}(\tilde{x}^{\alpha_0}) < 1$  via (9), we have by (11) and (10) that for all small enough  $\varepsilon > 0$ ,

$$\hat{\rho}(\tilde{x}^{\alpha_0} + \varepsilon \tilde{w}) < 1, \quad \tilde{x}^{\alpha_0} + \varepsilon \tilde{w} \in U.$$

Hence, (7) implies  $\lambda(\tilde{x}^{\alpha_0} + \varepsilon \tilde{w}) \leq \lambda(\tilde{z})$ . However, we have, by (8), an opposite inequality  $\lambda(\tilde{x}^{\alpha_0} + \varepsilon \tilde{w}) = \lambda(\tilde{x}^{\alpha_0}) + \varepsilon |\tilde{w}|^2 > \lambda(\tilde{x}^{\alpha_0}) = \lambda(\tilde{z})$ , leading to a contradiction. In summary, we have shown that

$$\lambda(\tilde{x}) < \lambda(\tilde{z}), \forall \tilde{x} \in U. \quad (12)$$

Since  $\hat{\rho}(0) = 0$ , we have  $0 \in U$ . Letting  $\tilde{x} = 0$  in (12) yields  $\lambda(\tilde{z}) > 0$ . So we can re-scale  $\lambda$  such that

$$\lambda(\tilde{z}) = 1 = \hat{\rho}(\tilde{z}).$$

Thus, (12) becomes

$$\lambda(\tilde{x}) < 1 \text{ for all } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x}) < 1,$$

from which (6) holds.  $\square$

**Lemma 2.** Let  $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \leq y_2 \leq \dots \leq y_n\}$ , and denote  $\mathcal{B}^o$  to be the interior of  $\mathcal{B}$ . For any fixed  $\tilde{z} = (z_1, \dots, z_n) \in \mathcal{B}^o$  and any  $\hat{\rho}$  satisfying Axioms D1-D4, there exists a weight  $\tilde{w} = (w_1, \dots, w_n)$  such that

$$\sum_{i=1}^n w_i = 1, \tag{13}$$

$$w_k \geq 0, k = 1, \dots, n, \tag{14}$$

$$\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^n w_i x_i, \text{ for } \forall \tilde{x} \in \mathcal{B}, \text{ and } \hat{\rho}(\tilde{z}) = \sum_{i=1}^n w_i z_i. \tag{15}$$

*Proof.* We will show this by considering three cases.

Case 1:  $\hat{\rho}(\tilde{z}) = 1$ .

From Lemma 1, there exists a weight  $\tilde{w} = (w_1, \dots, w_n)$  such that the linear functional  $\lambda(\tilde{x}) := \sum_{i=1}^n w_i x_i$  satisfies (5) and (6).

First we prove that  $\tilde{w}$  satisfies (13). For this, it is sufficient to show that  $\lambda(\mathbf{1}) = \sum_{i=1}^n w_i = 1$ . To this end, first note that for any  $c < 1$  Axiom D1 implies  $\hat{\rho}(c\mathbf{1}) = c < 1$ . Thus, (6) implies  $\lambda(c\mathbf{1}) < 1$ , and, by continuity of  $\lambda$ , we obtain that  $\lambda(\mathbf{1}) \leq 1$ . Secondly,  $\forall c > 1$  Axiom D1 implies  $\hat{\rho}(2\tilde{z} - c\mathbf{1}) = 2\hat{\rho}(\tilde{z}) - c = 2 - c < 1$ . Then it follows from (6) and (5) that  $1 > \lambda(2\tilde{z} - c\mathbf{1}) = 2\lambda(\tilde{z}) - c\lambda(\mathbf{1}) = 2 - c\lambda(\mathbf{1})$ , i.e.  $\lambda(\mathbf{1}) > 1/c$  for any  $c > 1$ . So  $\lambda(\mathbf{1}) \geq 1$ , and  $\tilde{w}$  satisfies (13).

Next, we will prove that  $\tilde{w}$  satisfies (14). Let  $\tilde{e}_k = (0, \dots, 0, 1, 0, \dots, 0)$  be the  $k$ -th standard basis of  $\mathbb{R}^n$ . Then  $w_k = \lambda(\tilde{e}_k)$ . Since  $\tilde{z} \in \mathcal{B}^o$ , there exists  $\delta > 0$  such that  $\tilde{z} - \delta\tilde{e}_k \in \mathcal{B}$ . For any  $\varepsilon > 0$ , we have

$$\hat{\rho}(\tilde{z} - \delta\tilde{e}_k - \varepsilon\mathbf{1}) = \hat{\rho}(\tilde{z} - \delta\tilde{e}_k) - \varepsilon \leq \hat{\rho}(\tilde{z}) - \varepsilon = 1 - \varepsilon < 1,$$

where the inequality follows from the monotonicity in Axiom D2. Then (6) and (5) implies

$$1 > \lambda(\tilde{z} - \delta\tilde{e}_k - \varepsilon\mathbf{1}) = \lambda(\tilde{z}) - \delta\lambda(\tilde{e}_k) - \varepsilon\lambda(\mathbf{1}) = 1 - \varepsilon - \delta\lambda(\tilde{e}_k).$$

Hence  $w_k = \lambda(\tilde{e}_k) > -\varepsilon/\delta$ , and the conclusion follows by letting  $\varepsilon$  goes to 0.

Finally, we will prove that  $\tilde{w}$  satisfies (15). It follows from Axiom D1 and (6) that

$$\forall c > 0, \quad \lambda(\tilde{x}) < c, \quad \text{for all } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x}) < c. \tag{16}$$

For any  $c \leq 0$ , we choose  $b > 0$  such that  $b + c > 0$ . Then by (16), we have

$$\lambda(\tilde{x} + b\mathbf{1}) < c + b, \quad \text{for all } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x} + b\mathbf{1}) < c + b.$$

Since  $\lambda(\tilde{x} + b\mathbf{1}) = \lambda(\tilde{x}) + b\lambda(\mathbf{1}) = \lambda(\tilde{x}) + b$  and  $\hat{\rho}(\tilde{x} + b\mathbf{1}) = \hat{\rho}(\tilde{x}) + b$  we have

$$\forall c \leq 0, \quad \lambda(\tilde{x}) < c, \quad \text{for all } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x}) < c. \quad (17)$$

It follows from (16) and (17) that

$$\hat{\rho}(\tilde{x}) \geq \lambda(\tilde{x}), \quad \text{for all } \tilde{x} \in \mathcal{B},$$

which in combination with  $\tilde{\rho}(\tilde{z}) = \lambda(\tilde{z}) = 1$  completes the proof of (15).

Case 2:  $\hat{\rho}(\tilde{z}) \neq 1$  and  $\hat{\rho}(\tilde{z}) > 0$ .

Since  $\hat{\rho}\left(\frac{1}{\hat{\rho}(\tilde{z})}\tilde{z}\right) = 1$  and  $\frac{1}{\hat{\rho}(\tilde{z})}\tilde{z}$  is still an interior point of  $\mathcal{B}$ , it follows from the result proved in Case 1 that there exists a linear functional  $\lambda(\tilde{x}) := \sum_{i=1}^n w_i x_i$ , with  $\tilde{w} = (w_1, \dots, w_n)$  satisfying (13), (14) and

$$\hat{\rho}(\tilde{x}) \geq \lambda(\tilde{x}), \forall \tilde{x} \in \mathcal{B}, \text{ and } \hat{\rho}\left(\frac{1}{\hat{\rho}(\tilde{z})}\tilde{z}\right) = \lambda\left(\frac{1}{\hat{\rho}(\tilde{z})}\tilde{z}\right),$$

or equivalently

$$\hat{\rho}(\tilde{x}) \geq \lambda(\tilde{x}), \forall \tilde{x} \in \mathcal{B}, \text{ and } \hat{\rho}(\tilde{z}) = \lambda(\tilde{z}).$$

Thus,  $\tilde{w}$  also satisfies (15).

Case 3:  $\hat{\rho}(\tilde{z}) \leq 0$ .

Choose  $b > 0$  such that  $\hat{\rho}(\tilde{z} + b\mathbf{1}) > 0$ . Since  $\tilde{z} + b\mathbf{1}$  is an interior point of  $\mathcal{B}$ , it follows from the result proved in Case 2 that there exists a linear functional  $\lambda(\tilde{x}) := \sum_{i=1}^n w_i x_i$  with  $\tilde{w} = (w_1, \dots, w_n)$  satisfying (13), (14) and

$$\hat{\rho}(\tilde{x}) \geq \lambda(\tilde{x}), \forall \tilde{x} \in \mathcal{B}, \text{ and } \hat{\rho}(\tilde{z} + b\mathbf{1}) = \lambda(\tilde{z} + b\mathbf{1}),$$

or equivalently

$$\hat{\rho}(\tilde{x}) \geq \lambda(\tilde{x}), \forall \tilde{x} \in \mathcal{B}, \text{ and } \hat{\rho}(\tilde{z}) = \lambda(\tilde{z}).$$

Thus,  $\tilde{w}$  also satisfies (15).  $\square$

*Proof of Theorem 1.* (I) The proof of the “if” part. Suppose  $\hat{\rho}$  is defined by (4). Then obviously  $\hat{\rho}$  satisfies Axioms D1 and D4.

To check Axiom D2, write

$$(y_{(1)}, y_{(2)}, \dots, y_{(n)}) = (y_{i_1}, y_{i_2}, \dots, y_{i_n}),$$

where  $(i_1, \dots, i_n)$  is a permutation of  $(1, \dots, n)$ . Then for any  $\tilde{x} \leq \tilde{y}$ , we have

$$y_{(k)} \geq \max\{y_{i_j}, j = 1, \dots, k\} \geq \max\{x_{i_j}, j = 1, \dots, k\} \geq x_{(k)}, \quad 1 \leq k \leq n,$$

which implies that  $\hat{\rho}$  satisfies Axiom D2 because

$$\hat{\rho}(\tilde{y}) = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i y_{(i)} \right\} \geq \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_{(i)} \right\} = \hat{\rho}(\tilde{x}).$$

To check Axiom D3, note that if  $\tilde{x}$  and  $\tilde{y}$  are comonotonic, then there exists a permutation  $(i_1, \dots, i_n)$  of  $(1, \dots, n)$  such that  $x_{i_1} \leq x_{i_2} \leq \dots \leq x_{i_n}$  and  $y_{i_1} \leq y_{i_2} \leq \dots \leq y_{i_n}$ . Hence, we have  $(\tilde{x} + \tilde{y})_{(i)} = x_{(i)} + y_{(i)}$ ,  $i = 1, \dots, n$ . Therefore,

$$\begin{aligned} \hat{\rho}(\tilde{x} + \tilde{y}) &= \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i (\tilde{x} + \tilde{y})_{(i)} \right\} = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i (x_{(i)} + y_{(i)}) \right\} \\ &\leq \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_{(i)} \right\} + \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i y_{(i)} \right\} = \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y}), \end{aligned}$$

which implies that  $\hat{\rho}$  satisfies Axiom D3.

(II) The proof of the “only if” part. By Axiom D4, we only need to show that there exists a set of weights  $\mathcal{W} = \{\tilde{w} = (w_1, \dots, w_n)\} \subset \mathbb{R}^n$  with each  $\tilde{w} \in \mathcal{W}$  satisfies  $\sum_{i=1}^n w_i = 1$  and  $w_i \geq 0$ ,  $\forall 1 \leq i \leq n$ , such that

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_i \right\}, \quad \forall \tilde{x} \in \mathcal{B}, \quad (18)$$

where recall that  $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \leq y_2 \leq \dots \leq y_n\}$ .

By Lemma 2, for any point  $\tilde{y} \in \mathcal{B}^\circ$ , there exists a weight  $\tilde{w}^{(\tilde{y})}$  satisfying (13), (14) and (15). Therefore, we can take the collection of such weights as

$$\mathcal{W} = \{\tilde{w}^{(\tilde{y})} \mid \tilde{y} \in \mathcal{B}^\circ\}.$$

Then from (15), for any fixed  $\tilde{x} \in \mathcal{B}^\circ$  we have

$$\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^n w_i^{(\tilde{y})} x_i, \quad \forall \tilde{y} \in \mathcal{B}^\circ,$$

$$\hat{\rho}(\tilde{x}) = \sum_{i=1}^n w_i^{(\tilde{x})} x_i,$$

Therefore,

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{y} \in \mathcal{B}^\circ} \left\{ \sum_{i=1}^n w_i^{(\tilde{y})} x_i \right\} = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_i \right\}, \quad \forall \tilde{x} \in \mathcal{B}^\circ, \quad (19)$$

where each  $\tilde{w} \in \mathcal{W}$  satisfies (13) and (14).

Next, we will prove that the above equality is also true for any boundary point of  $\mathcal{B}$ , i.e.,

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_i \right\}, \quad \forall \tilde{x} \in \partial \mathcal{B}. \quad (20)$$

Let  $\tilde{x}^0$  be any boundary point of  $\mathcal{B}$ . Then there exists a sequence  $\{\tilde{x}^k\}_{k=1}^\infty \subset \mathcal{B}^\circ$  such that  $\tilde{x}^k \rightarrow \tilde{x}^0$  as  $k \rightarrow \infty$ . By the continuity of  $\hat{\rho}$ , we have

$$\hat{\rho}(\tilde{x}^0) = \lim_{k \rightarrow \infty} \hat{\rho}(\tilde{x}^k) = \lim_{k \rightarrow \infty} \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_i^k \right\}, \quad (21)$$

where the last equality follows from (19). If we can interchange sup and limit in (21), i.e. if

$$\lim_{k \rightarrow \infty} \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_i^k \right\} = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \lim_{k \rightarrow \infty} \sum_{i=1}^n w_i x_i^k \right\} = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_i^0 \right\}, \quad (22)$$

then (20) holds and the proof is complete.

To show (22), note that we have by Cauchy-Schwartz inequality

$$\left| \sum_{i=1}^n w_i x_i^k - \sum_{i=1}^n w_i x_i^0 \right| \leq \left( \sum_{i=1}^n (w_i)^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^n (x_i^k - x_i^0)^2 \right)^{\frac{1}{2}} \leq \left( \sum_{i=1}^n (x_i^k - x_i^0)^2 \right)^{\frac{1}{2}}, \quad \forall \tilde{w} \in \mathcal{W},$$

because  $w_i \geq 0$  and  $\sum_{i=1}^n w_i = 1, \forall \tilde{w} \in \mathcal{W}$ . Therefore,  $\sum_{i=1}^n w_i x_i^k \rightarrow \sum_{i=1}^n w_i x_i^0$  uniformly for all  $\tilde{w} \in \mathcal{W}$  and (22) follows.  $\square$

## 5 Another Representation of the Natural Risk Statistic via Acceptance Sets

Similar to the coherent risk measure, the proposed natural risk statistic can also be characterized via acceptance sets. A *statistical acceptance set* is a subset of  $\mathbb{R}^n$ . Given a statistical acceptance set  $\mathcal{A} \in \mathbb{R}^n$ , the risk statistic  $\hat{\rho}_{\mathcal{A}}$  associated with  $\mathcal{A}$  is defined to be the minimal amount of risk-free investment that has to be added to the original position so that the resulting position is acceptable, or in mathematical form

$$\hat{\rho}_{\mathcal{A}}(\tilde{x}) = \inf \{ m \mid \tilde{x} - m \mathbf{1} \in \mathcal{A} \}, \quad \forall \tilde{x} \in \mathbb{R}^n. \quad (23)$$

On the other hand, given a risk statistic  $\hat{\rho}$ , one can define the statistical acceptance set associated with  $\hat{\rho}$  by

$$\mathcal{A}_{\hat{\rho}} = \{ \tilde{x} \in \mathbb{R}^n \mid \hat{\rho}(\tilde{x}) \leq 0 \}. \quad (24)$$

We shall postulate the following axioms for statistical acceptance set  $\mathcal{A}$ :

**Axiom E1.** The acceptance set  $\mathcal{A}$  contains  $\mathbb{R}_-^n$  where  $\mathbb{R}_-^n = \{\tilde{x} \in \mathbb{R}^n \mid x_i \leq 0, i = 1, \dots, n\}$ .

**Axiom E2.** The acceptance set  $\mathcal{A}$  does not intersect the set  $\mathbb{R}_{++}^n$  where  $\mathbb{R}_{++}^n = \{\tilde{x} \in \mathbb{R}^n \mid x_i > 0, i = 1, \dots, n\}$ .

**Axiom E3.** If  $\tilde{x}$  and  $\tilde{y}$  are comonotonic and  $\tilde{x} \in \mathcal{A}$ ,  $\tilde{y} \in \mathcal{A}$ , then  $\lambda\tilde{x} + (1 - \lambda)\tilde{y} \in \mathcal{A}$ , for  $\forall \lambda \in [0, 1]$ .

**Axiom E4.** The acceptance set  $\mathcal{A}$  is positively homogeneous, i.e., if  $\tilde{x} \in \mathcal{A}$ , then  $\lambda\tilde{x} \in \mathcal{A}$  for all  $\lambda \geq 0$ .

**Axiom E5.** If  $\tilde{x} \leq \tilde{y}$  and  $\tilde{y} \in \mathcal{A}$ , then  $\tilde{x} \in \mathcal{A}$ .

**Axiom E6.** If  $\tilde{x} \in \mathcal{A}$ , then  $(x_{i_1}, \dots, x_{i_n}) \in \mathcal{A}$  for any permutation  $(i_1, \dots, i_n)$ .

We will show that a natural risk statistic and a statistical acceptance set satisfying axioms C.1-C.6 are mutually representable. More precisely, we have the following Theorem:

**Theorem 2.** (I) If  $\hat{\rho}$  is a natural risk statistic, then the statistical acceptance set  $\mathcal{A}_{\hat{\rho}}$  is closed and satisfies axioms E1-E6.

(II) If a statistical acceptance set  $\mathcal{A}$  satisfies axioms E1-E6, then the risk statistic  $\hat{\rho}_{\mathcal{A}}$  is a natural risk statistic.

(III) If  $\hat{\rho}$  is a natural risk statistic, then  $\hat{\rho} = \hat{\rho}_{\mathcal{A}_{\hat{\rho}}}$ .

(IV) If a statistical acceptance set  $\mathcal{D}$  satisfies axioms E1-E6, then  $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}} = \bar{\mathcal{D}}$ , the closure of  $\mathcal{D}$ .

**Remark:** Theorem 2 shows that the risk statistic  $\hat{\rho}$  calculated from the data  $\tilde{x}$  is equivalent to the amount of risk-free investment that has to be added to make the original position acceptable. This alternative characterization of the natural risk statistic is consistent with a similar characterization of the coherent risk measure in Artzner et al. ??.

*Proof.* (I) (1) For  $\forall \tilde{x} \leq 0$ , Axiom D2 implies  $\hat{\rho}(\tilde{x}) \leq \hat{\rho}(0) = 0$ , hence  $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$  by definition. Thus, E1 holds. (2) For any  $\tilde{x} \in \mathbb{R}_{++}^n$ , there exists  $\alpha > 0$  such that  $0 \leq \tilde{x} - \alpha\mathbf{1}$ . Axioms D2 and D1 imply that  $\hat{\rho}(0) \leq \hat{\rho}(\tilde{x} - \alpha\mathbf{1}) = \hat{\rho}(\tilde{x}) - \alpha$ . So  $\hat{\rho}(\tilde{x}) \geq \alpha > 0$  and henceforth  $\tilde{x} \notin \mathcal{A}_{\hat{\rho}}$ , i.e., E2 holds. (3) If  $\tilde{x}$  and  $\tilde{y}$  are comonotonic and  $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$ ,  $\tilde{y} \in \mathcal{A}_{\hat{\rho}}$ , then  $\hat{\rho}(\tilde{x}) \leq 0$ ,  $\hat{\rho}(\tilde{y}) \leq 0$ , and  $\lambda\tilde{x}$  and  $(1 - \lambda)\tilde{y}$  are comonotonic for any  $\lambda \in [0, 1]$ . Thus D3 implies

$$\hat{\rho}(\lambda\tilde{x} + (1 - \lambda)\tilde{y}) \leq \hat{\rho}(\lambda\tilde{x}) + \hat{\rho}((1 - \lambda)\tilde{y}) = \lambda\hat{\rho}(\tilde{x}) + (1 - \lambda)\hat{\rho}(\tilde{y}) \leq 0.$$

Hence  $\lambda\tilde{x} + (1 - \lambda)\tilde{y} \in \mathcal{A}_{\hat{\rho}}$ , i.e., E3 holds. (4) For any  $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$  and  $a > 0$ , we have  $\hat{\rho}(\tilde{x}) \leq 0$  and D1 implies  $\hat{\rho}(a\tilde{x}) = a\hat{\rho}(\tilde{x}) \leq 0$ . Thus,  $a\tilde{x} \in \mathcal{A}_{\hat{\rho}}$ , i.e., E4 holds. (5) For any  $\tilde{x} \leq \tilde{y}$  and  $\tilde{y} \in \mathcal{A}_{\hat{\rho}}$ , we have  $\hat{\rho}(\tilde{y}) \leq 0$ . By D2,  $\hat{\rho}(\tilde{x}) \leq \hat{\rho}(\tilde{y}) \leq 0$ . Hence  $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$ , i.e., E5 holds. (6) If  $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$ ,

then  $\hat{\rho}(\tilde{x}) \leq 0$ . For any permutation  $(i_1, \dots, i_n)$ , D4 implies  $\hat{\rho}((x_{i_1}, \dots, x_{i_n})) = \hat{\rho}(\tilde{x}) \leq 0$ . So  $(x_{i_1}, \dots, x_{i_n}) \in \mathcal{A}_{\hat{\rho}}$ , i.e., E6 holds. (7) Suppose  $\tilde{x}^k \in \mathcal{A}_{\hat{\rho}}, k = 1, 2, \dots$ , and  $\tilde{x}^k \rightarrow \tilde{x}$  as  $k \rightarrow \infty$ . Then  $\hat{\rho}(\tilde{x}^k) \leq 0, \forall k$ . Suppose the limit  $\tilde{x} \notin \mathcal{A}_{\hat{\rho}}$ . Then  $\hat{\rho}(\tilde{x}) > 0$ . There exists  $\delta > 0$  such that  $\hat{\rho}(\tilde{x} - \delta \mathbf{1}) > 0$ . Since  $\tilde{x}^k \rightarrow \tilde{x}$ , it follows that there exists  $K \in \mathbb{N}$  such that  $\tilde{x}^K > \tilde{x} - \delta \mathbf{1}$ . By D2,  $\hat{\rho}(\tilde{x}^K) \geq \hat{\rho}(\tilde{x} - \delta \mathbf{1}) > 0$ , which contradicts to  $\hat{\rho}(\tilde{x}^K) \leq 0$ . So  $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$ , i.e.,  $\mathcal{A}_{\hat{\rho}}$  is closed.

(II) (1) For  $\forall \tilde{x} \in \mathbb{R}^n, \forall b \in \mathbb{R}$ , we have

$$\hat{\rho}_{\mathcal{A}}(\tilde{x} + b\mathbf{1}) = \inf\{m \mid \tilde{x} + b\mathbf{1} - m\mathbf{1} \in \mathcal{A}\} = b + \inf\{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{A}\} = b + \hat{\rho}_{\mathcal{A}}(\tilde{x}).$$

For  $\forall \tilde{x} \in \mathbb{R}^n, \forall a \geq 0$ , if  $a = 0$ , then

$$\hat{\rho}_{\mathcal{A}}(a\tilde{x}) = \inf\{m \mid 0 - m\mathbf{1} \in \mathcal{A}\} = 0 = a \cdot \hat{\rho}_{\mathcal{A}}(\tilde{x}),$$

where the second equality follows from E1 and E2. If  $a > 0$ , then

$$\begin{aligned} \hat{\rho}_{\mathcal{A}}(a\tilde{x}) &= \inf\{m \mid a\tilde{x} - m\mathbf{1} \in \mathcal{A}\} = a \cdot \inf\{u \mid a(\tilde{x} - u\mathbf{1}) \in \mathcal{A}\} \\ &= a \cdot \inf\{u \mid \tilde{x} - u\mathbf{1} \in \mathcal{A}\} = a \cdot \hat{\rho}_{\mathcal{A}}(\tilde{x}), \end{aligned}$$

by E4. Therefore, D1 holds. (2) Suppose  $\tilde{x} \leq \tilde{y}$ . For any  $m \in \mathbb{R}$ , if  $\tilde{y} - m\mathbf{1} \in \mathcal{A}$ , then E5 and  $\tilde{x} - m\mathbf{1} \leq \tilde{y} - m\mathbf{1}$  imply that  $\tilde{x} - m\mathbf{1} \in \mathcal{A}$ . Hence  $\{m \mid \tilde{y} - m\mathbf{1} \in \mathcal{A}\} \subset \{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{A}\}$ . By taking infimum on both sides, we obtain  $\hat{\rho}_{\mathcal{A}}(\tilde{y}) \geq \hat{\rho}_{\mathcal{A}}(\tilde{x})$ , i.e., D2 holds. (3) Suppose  $\tilde{x}$  and  $\tilde{y}$  are comonotonic. For any  $m$  and  $n$  such that  $\tilde{x} - m\mathbf{1} \in \mathcal{A}, \tilde{y} - n\mathbf{1} \in \mathcal{A}$ , since  $\tilde{x} - m\mathbf{1}$  and  $\tilde{y} - n\mathbf{1}$  are comonotonic, it follows from E3 that  $\frac{1}{2}(\tilde{x} - m\mathbf{1}) + \frac{1}{2}(\tilde{y} - n\mathbf{1}) \in \mathcal{A}$ . By E4, the previous formula implies  $\tilde{x} + \tilde{y} - (m + n)\mathbf{1} \in \mathcal{A}$ . Therefore,

$$\hat{\rho}_{\mathcal{A}}(\tilde{x} + \tilde{y}) \leq m + n.$$

Taking infimum of all  $m$  and  $n$  satisfying  $\tilde{x} - m\mathbf{1} \in \mathcal{A}, \tilde{y} - n\mathbf{1} \in \mathcal{A}$ , on both sides of above inequality yields

$$\hat{\rho}_{\mathcal{A}}(\tilde{x} + \tilde{y}) \leq \hat{\rho}_{\mathcal{A}}(\tilde{x}) + \hat{\rho}_{\mathcal{A}}(\tilde{y}).$$

So D3 holds. (4) Fix any  $\tilde{x} \in \mathbb{R}^n$  and any permutation  $(i_1, \dots, i_n)$ . Then for any  $m \in \mathbb{R}$ , E6 implies that  $\tilde{x} - m\mathbf{1} \in \mathcal{A}$  if and only if  $(x_{i_1}, \dots, x_{i_n}) - m\mathbf{1} \in \mathcal{A}$ . Hence

$$\{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{A}\} = \{m \mid (x_{i_1}, \dots, x_{i_n}) - m\mathbf{1} \in \mathcal{A}\}.$$

Taking infimum on both sides, we obtain  $\hat{\rho}_{\mathcal{A}}(\tilde{x}) = \hat{\rho}_{\mathcal{A}}((x_{i_1}, \dots, x_{i_n}))$ , i.e., D4 holds.

(III) For  $\forall \tilde{x} \in \mathbb{R}^n$ , we have

$$\hat{\rho}_{\mathcal{A}_{\hat{\rho}}}(\tilde{x}) = \inf\{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{A}_{\hat{\rho}}\} = \inf\{m \mid \hat{\rho}(\tilde{x} - m\mathbf{1}) \leq 0\} = \inf\{m \mid \hat{\rho}(\tilde{x}) \leq m\} \geq \hat{\rho}(\tilde{x}),$$

via D1. On the other hand, for any  $\delta > \hat{\rho}(\tilde{x})$ , we have

$$\begin{aligned} \delta > \hat{\rho}(\tilde{x}) &\Rightarrow \hat{\rho}(\tilde{x} - \delta\mathbf{1}) < 0 \text{ (since } \hat{\rho} \text{ satisfies D1)} \\ &\Rightarrow \tilde{x} - \delta\mathbf{1} \in \mathcal{A}_{\hat{\rho}} \text{ (by definition)} \\ &\Rightarrow \hat{\rho}_{\mathcal{A}_{\hat{\rho}}}(\tilde{x} - \delta\mathbf{1}) \leq 0 \text{ (by definition)} \\ &\Rightarrow \hat{\rho}_{\mathcal{A}_{\hat{\rho}}}(\tilde{x}) \leq \delta \text{ (since } \hat{\rho}_{\mathcal{A}_{\hat{\rho}}} \text{ satisfies D1)} \end{aligned}$$

Letting  $\delta \downarrow \hat{\rho}(\tilde{x})$ , we obtain  $\hat{\rho}_{\mathcal{A}_{\hat{\rho}}}(\tilde{x}) \leq \hat{\rho}(\tilde{x})$ . Therefore,  $\hat{\rho}(\tilde{x}) = \hat{\rho}_{\mathcal{A}_{\hat{\rho}}}(\tilde{x})$ .

(IV) For any  $\tilde{x} \in \mathcal{D}$ , we have  $\hat{\rho}_{\mathcal{D}}(\tilde{x}) \leq 0$ . Hence  $\tilde{x} \in \mathcal{A}_{\hat{\rho}_{\mathcal{D}}}$ . Therefore,  $\mathcal{D} \subset \mathcal{A}_{\hat{\rho}_{\mathcal{D}}}$ . By the results (I) and (II),  $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}}$  is closed. So  $\bar{\mathcal{D}} \subset \mathcal{A}_{\hat{\rho}_{\mathcal{D}}}$ . On the other hand, for any  $\tilde{x} \in \mathcal{A}_{\hat{\rho}_{\mathcal{D}}}$ , we have by definition that  $\hat{\rho}_{\mathcal{D}}(\tilde{x}) \leq 0$ , i.e.,  $\inf\{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{D}\} \leq 0$ . If  $\inf\{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{D}\} < 0$ , then there exists  $m < 0$  such that  $\tilde{x} - m\mathbf{1} \in \mathcal{D}$ . Then since  $\tilde{x} < \tilde{x} - m\mathbf{1}$  by E5  $\tilde{x} \in \mathcal{D}$ . Otherwise,  $\inf\{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{D}\} = 0$ . Then there exists  $m_k$  such that  $m_k \downarrow 0$  as  $k \rightarrow \infty$  and  $\tilde{x} - m_k\mathbf{1} \in \mathcal{D}$ . Hence  $\tilde{x} \in \bar{\mathcal{D}}$ . In either case we obtain that  $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}} \subset \bar{\mathcal{D}}$ . Therefore, we conclude that  $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}} = \bar{\mathcal{D}}$ .  $\square$

## 6 Relation between the Natural Risk Statistic, Coherent Risk Measure and Insurance Risk Measure

### 6.1 Relation with Coherent Risk Measures

To compare natural risk statistics with coherent risk measure in a formal manner, we first have to extend the coherence risk measure to coherent risk statistics.

**Definition 2.** A risk statistic  $\hat{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}$  is called a coherent risk statistic, if it satisfies Axioms D1, D2 and the following axiom F3:

**Axiom F3.** Subadditivity:  $\hat{\rho}(\tilde{x} + \tilde{y}) \leq \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y})$ , for every  $\tilde{x}, \tilde{y} \in \mathbb{R}^n$ .

By using the result from Huber [18], one can easily see that a risk statistic is a coherent risk statistic if and only if there exists a set of weights  $\mathcal{W} = \{\tilde{w} = (w_1, \dots, w_n)\} \subset \mathbb{R}^n$  with each  $\tilde{w} \in \mathcal{W}$  satisfying  $\sum_{i=1}^n w_i = 1$  and  $w_i \geq 0, \forall 1 \leq i \leq n$ , such that

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_i \right\}, \quad \forall \tilde{x} \in \mathbb{R}^n. \quad (25)$$

The next theorem shows the connection between natural risk statistics and coherent risk statistics.

**Theorem 3.** Consider a fixed scenario set  $\mathcal{W}$ , where each  $\tilde{w} \in \mathcal{W}$  satisfies  $\sum_{i=1}^n w_i = 1$  and  $w_i \geq 0, \forall 1 \leq i \leq n$ . Let  $\hat{\rho}$  be a natural risk statistic induced by  $\mathcal{W}$ :

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_{(i)} \right\}, \quad \forall \tilde{x} \in \mathbb{R}^n, \quad (26)$$

If every weight  $\tilde{w}$  is monotonic, i.e.,

$$w_1 \leq w_2 \leq \dots \leq w_n, \quad \forall \tilde{w} \in \mathcal{W}, \quad (27)$$

then  $\hat{\rho}$  satisfies subadditivity and is, therefore, a coherent risk statistic.

**Remark:** Comparing (25), (26), and (27), we see the main differences between the natural risk statistic and the coherent risk statistic:

(1) A natural risk statistic is a supremum of L-statistic (which is a weighted average of order statistics), while a coherent risk statistic is a supremum of weighted sample average. There is no simple linear function that can transform L-statistic to weighted sample average.

(2) Although VaR is not a coherent risk statistic, VaR is a natural risk statistic. In other words, though being simple, VaR is not without justification, as it also satisfies a different set of axioms.

(3) If one assigns larger weights to larger observations, resulting in (27), then a natural risk statistic become a coherent risk statistic. However, assigning larger weights to larger observations leads to less robust risk statistics, as the statistics become more sensitive to larger observations.

*Proof of Theorem 3.* We only need to show that under condition (27), the risk statistic (26) satisfies subadditivity for any  $\tilde{x}$  and  $\tilde{y} \in \mathbb{R}^n$ . Let  $(k_1, \dots, k_n)$  be the permutation of  $(1, \dots, n)$  such that  $(\tilde{x} + \tilde{y})_{k_1} \leq (\tilde{x} + \tilde{y})_{k_2} \leq \dots \leq (\tilde{x} + \tilde{y})_{k_n}$ . Then for  $i = 1, \dots, n-1$ , the partial sum up to  $i$  satisfies

$$\sum_{j=1}^i (\tilde{x} + \tilde{y})_{(j)} = \sum_{j=1}^i (\tilde{x} + \tilde{y})_{k_j} = \sum_{j=1}^i (x_{k_j} + y_{k_j}) \geq \sum_{j=1}^i (x_{(j)} + y_{(j)}). \quad (28)$$

In addition, we have for the total sum

$$\sum_{j=1}^n (\tilde{x} + \tilde{y})_{(j)} = \sum_{j=1}^n (x_j + y_j) = \sum_{j=1}^n (x_{(j)} + y_{(j)}). \quad (29)$$

Re-arranging the summation terms yields

$$\hat{\rho}(\tilde{x} + \tilde{y}) = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i (\tilde{x} + \tilde{y})_{(i)} \right\} = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^{n-1} (w_i - w_{i+1}) \sum_{j=1}^i (\tilde{x} + \tilde{y})_{(j)} + w_n \sum_{j=1}^n (\tilde{x} + \tilde{y})_{(j)} \right\},$$

This, along with the fact  $w_i - w_{i+1} \leq 0$  and equations (28) and (29), shows that

$$\begin{aligned} \hat{\rho}(\tilde{x} + \tilde{y}) &\leq \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^{n-1} (w_i - w_{i+1}) \sum_{j=1}^i (x_{(j)} + y_{(j)}) + w_n \sum_{j=1}^n (x_{(j)} + y_{(j)}) \right\} \\ &= \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_{(i)} + \sum_{i=1}^n w_i y_{(i)} \right\} \\ &\leq \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i x_{(i)} \right\} + \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^n w_i y_{(i)} \right\} = \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y}), \end{aligned}$$

from which the proof is completed.  $\square$

## 6.2 Relation with the Insurance Risk Measure

Similar to the coherent risk statistic, we can extend the insurance risk measure to the insurance risk statistic as follows:

**Definition 3.** A risk statistic  $\hat{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}$  is called a insurance risk statistic, if it satisfies the following axioms G1-G3:

**Axiom G1.** Monotonicity:  $\hat{\rho}(\tilde{x}) \leq \hat{\rho}(\tilde{y})$ , if  $\tilde{x} \leq \tilde{y}$ .

**Axiom G2.** Comonotonic additivity:  $\hat{\rho}(\tilde{x} + \tilde{y}) = \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y})$ , if  $\tilde{x}$  and  $\tilde{y}$  are comonotonic.

**Axiom G3.** Continuity:

$$\lim_{d \rightarrow 0} \hat{\rho}((\tilde{x} - d)^+) = \hat{\rho}(\tilde{x}^+), \quad \lim_{d \rightarrow \infty} \hat{\rho}(\min(\tilde{x}, d)) = \hat{\rho}(\tilde{x}), \quad \lim_{d \rightarrow -\infty} \hat{\rho}(\max(\tilde{x}, d)) = \hat{\rho}(\tilde{x}),$$

where  $(\tilde{x} - d)^+ = \max(\tilde{x} - d, 0)$ .

Given an observation  $\tilde{x} = (x_1, \dots, x_n)$  from the loss variable  $X$ , the empirical distribution of  $X$  is given by  $\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x \leq x_i\}}$ . Replacing the tail probability  $P(X > t)$  by empirical tail probability  $1 - \hat{F}(t)$  in the distorted probability representation of the insurance risk measure (2), we obtain the representation of insurance risk statistic

$$\hat{\rho}(\tilde{x}) = \int_{-\infty}^0 (g(1 - \hat{F}(t)) - 1) dt + \int_0^{\infty} g(1 - \hat{F}(t)) dt = \sum_{i=1}^n w_i x_{(i)}, \quad (30)$$

where  $w_i = g\left(\frac{n-i+1}{n}\right) - g\left(\frac{n-i}{n}\right)$ ,  $i = 1, \dots, n$ . Since  $g$  is nondecreasing,  $g(0) = 0$  and  $g(1) = 1$ , it follows that  $w_i \geq 0$ ,  $i = 1, \dots, n$ , and  $\sum_{i=1}^n w_i = 1$ .

Comparing (4) and (30), we see that the natural risk statistic is the supremum of L-statistics, while the insurance risk statistic is just one L-statistic. Therefore, the insurance risk statistics cannot incorporate different scenarios. On the other hand, each weight  $\tilde{w} = (w_1, \dots, w_n)$  in a natural risk statistic can be considered as a “scenario” in which (subjective or objective) evaluation of the importance of each ordered observations is specified. Hence, the nature risk statistic reflects the idea of evaluating the risk under different scenarios, similar to the coherent risk measure.

The following counterexample shows that if one incorporates different scenarios, then the comonotonic additivity may not hold, as the strict comonotonic subadditivity may prevail.

**A Counterexample:** Consider a natural risk statistic defined by

$$\hat{\rho}(\tilde{x}) = \max(0.5x_{(1)} + 0.5x_{(1)}, 0.72x_{(1)} + 0.08x_{(2)} + 0.2x_{(3)}), \quad \forall \tilde{x} \in \mathbb{R}^3.$$

Let  $\tilde{z} = (3, 2, 4)$  and  $\tilde{y} = (9, 4, 16)$ . By simple calculation we have

$$\hat{\rho}(\tilde{z} + \tilde{y}) = 9.28 < \hat{\rho}(\tilde{z}) + \hat{\rho}(\tilde{y}) = 2.5 + 6.8 = 9.3,$$

even though  $\tilde{x}$  and  $\tilde{y}$  are comonotonic. Therefore, the comonotonic additivity fails, and this natural risk statistic is not an insurance risk statistic. In summary, insurance risk statistic cannot incorporate those two simple scenarios.

## 7 Tail Conditional Median: a Robust Natural Risk Statistic

In this section, we propose a special case of the natural risk statistic, which we call the tail conditional median and compare it with an existing coherent risk measure, the tail conditional mean. Theoretical and numerical results are provided to illustrate the robustness of the proposed tail conditional median.

### 7.1 The Differences between Tail Conditional Expectation and Tail Conditional Median

A special case of coherent risk measures, the tail conditional expectation, or TCE for short, has gained popularity since it was proposed by Artzner et al.[3]. TCE satisfies subadditivity for continuous random variables, and also for discrete random variables if one define quantiles for discrete random variables properly; see Acerbi and Tasche ???. The TCE is also called expected shortfall by Acerbi et al.[1] and conditional value-at-risk by Rockafellar and Uryasev [23] and

Pflug [20]. More precisely, the TCE at level  $\alpha$  is defined by

$$\text{TCE}_\alpha(X) = \text{mean of the } \alpha\text{-tail distribution of } X, \quad (31)$$

where the distribution in question is one with distribution function  $G_\alpha(x)$  defined by

$$G_\alpha(x) = \begin{cases} 0 & \text{for } x < \text{VaR}_\alpha(X) \\ \frac{P(X \leq x) - \alpha}{1 - \alpha} & \text{for } x \geq \text{VaR}_\alpha(X). \end{cases}$$

If the distribution of  $X$  is continuous,

$$\text{TCE}_\alpha(X) = E[X|X \geq \text{VaR}_\alpha(X)]. \quad (32)$$

Essentially,  $\text{TCE}_\alpha(X)$  is the regularized version of the tail conditional expectation  $E[X|X \geq \text{VaR}_\alpha(X)]$ .

As we will see that TCE is not robust and is sensitive to model assumptions and outliers, here we propose an alternative, the tail conditional median (TCM), as a way of measuring risk to ameliorate the problem of robustness. The TCM at level  $\alpha$  is defined as

$$\text{TCM}_\alpha(X) = \text{median}[X|X \geq \text{VaR}_\alpha(X)]. \quad (33)$$

Essentially,  $\text{TCM}_\alpha(X)$  is the conditional median of  $X$  given that  $X \geq \text{VaR}_\alpha(X)$ . If  $X$  is continuous then  $\text{TCM}_\alpha(X) = \text{VaR}_{\frac{1+\alpha}{2}}(X)$ . However, for discrete random variable or data, one simply uses the definition (33) and there is a difference between  $\text{TCM}_\alpha(X)$  and  $\text{VaR}_{\frac{1+\alpha}{2}}(X)$ , depending on ways of defining quantiles for discrete random variables and discrete data.

There are several differences between the TCE and TCM. First, there are theoretical differences. For example, the TCM will not in general satisfy subadditivity, although TCE generally does; and the TCM is a natural risk statistic while the TCE is a coherent risk statistic.

Secondly, there may be significant numerical differences between TCE and TCM. To illustrate this, we shall use three data sets. The first one is a data set of auto insurance claims with sample size 736. In Table 1, we calculated the risk measure  $\text{TCE}_\alpha$  and  $\text{TCM}_\alpha$  with  $\alpha$  ranging from 95% to 99%. The absolute difference and the relative difference of the two measures are reported in the last two columns of the table, which shows that the relative differences can be as high as over 60% when  $\alpha < 99\%$  and about 30% when  $\alpha = 99\%$ . The second one is a data set of incendie insurance claims with sample size 736. Table 2 shows the  $\text{TCE}_\alpha$  and  $\text{TCM}_\alpha$  for this data set. The relative differences between  $\text{TCE}_\alpha$  and  $\text{TCM}_\alpha$  are also very significant for this data set .

Table 1: TCM and TCE for auto insurance claim

$\alpha$	$TCE_\alpha$	$TCM_\alpha$	$TCE_\alpha - TCM_\alpha$	$\frac{TCE_\alpha - TCM_\alpha}{TCE_\alpha}$
99.0%	6390627.0523	4489416.3847	1901210.6676	29.75%
98.5%	4454513.7015	1682970.0123	2771543.6892	62.22%
98.0%	3681944.0471	1384060.8997	2297883.1474	62.41%
97.5%	3014237.8755	1039186.8726	1975051.0028	65.52%
97.0%	2579508.4877	962778.2851	1616730.2026	62.68%
96.5%	2333814.6040	851033.8563	1482780.7477	63.53%
96.0%	2073066.4541	705136.3357	1367930.1185	65.99%
95.5%	1865231.5196	676514.4433	1188717.0763	63.73%
95.0%	1736077.5343	662045.2762	1074032.2581	61.87%

Table 2: TCM and TCE for incendie insurance claim

$\alpha$	$TCE_\alpha$	$TCM_\alpha$	$TCE_\alpha - TCM_\alpha$	$\frac{TCE_\alpha - TCM_\alpha}{TCE_\alpha}$
99.0%	326941430.7051	221019625.0743	105921805.6308	32.40%
98.5%	217524078.9696	68744819.4502	148779259.5194	68.40%
98.0%	174769416.7362	41636726.9290	133132689.8072	76.18%
97.5%	138375399.0228	24408413.1754	113966985.8474	82.36%
97.0%	114576698.9187	23633013.0148	90943685.9039	79.37%
96.5%	101683055.9714	18352824.0186	83330231.9528	81.95%
96.0%	88411141.0208	13845230.0549	74565910.9660	84.34%
95.5%	78274066.9951	10493594.7803	67780472.2148	86.59%
95.0%	72110217.3841	8572087.3993	63538129.9848	88.11%

The third data set is S&P 500 daily index from January 03, 1980 to December 21, 2005. More precisely, let  $\tilde{x} = (x_1, x_2, \dots, x_n)$  be the negative of the net daily return calculated from the data, where  $n = 6556$  here. Then we apply TCE and TCM to  $\tilde{x}$  to measure the risk. Table 7.1 shows the results for  $\alpha$  ranging from 95.0% to 99.9%. The relative differences of TCE and TCM are in the range of 12.72%  $\sim$  25.70%.

## 7.2 Robustness Comparison between the Tail Conditional Expectation and Tail Conditional Median

Next we show that the tail conditional median is more robust than the tail conditional expectation. The left panel of Figure 1 shows the value of  $TCE_\alpha$  with respect to  $\log(1 - \alpha)$  for Laplace distribution and T-distribution, where  $\alpha$  is in the range  $[0.95, 0.999]$ . As demonstrated in the graph, if the model assumes the loss distribution to be Laplace while the underlying true

$\alpha$	$TCE_\alpha$	$TCM_\alpha$	$TCE_\alpha - TCM_\alpha$	$\frac{TCE_\alpha - TCM_\alpha}{TCE_\alpha}$
99.9%	0.0922	0.0685	0.0237	25.70%
99.5%	0.0487	0.0389	0.0098	20.21%
99.0%	0.0383	0.0306	0.0078	20.24%
98.5%	0.0337	0.0280	0.0057	16.97%
98.0%	0.0308	0.0259	0.0050	16.15%
97.5%	0.0288	0.0245	0.0043	14.94%
97.0%	0.0272	0.0233	0.0038	14.13%
96.5%	0.0259	0.0224	0.0035	13.54%
96.0%	0.0248	0.0217	0.0032	12.72%
95.5%	0.0239	0.0207	0.0032	13.21%
95.0%	0.0231	0.0196	0.0035	15.05%

Table 3: The TCE and TCM for S&P 500 index daily losses (negative returns) from Jan 03, 1980 to Dec 21, 2005 with the range for  $\alpha$  is from 95.0% to 99.9%. The table shows a significant difference between the TCE and TCM.

loss distribution is T-distribution, the calculated TCE value can be far from the true value.

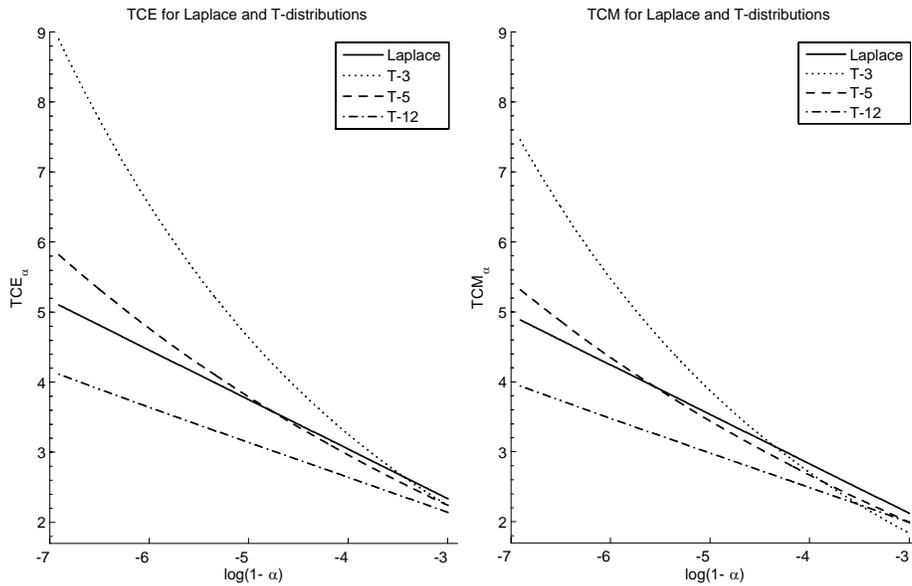


Figure 1: TCE and TCM for Laplace and T-distributions with degree of freedom 3, 5, 12. All distributions are normalized with mean 0 and variance 1. The TCM is less sensitive to changes in distribution, as the right panel has narrower range in y-axis.

The right panel of Figure 1 shows the value of  $TCM_\alpha$  with respect to  $\log(1 - \alpha)$  for Laplace distribution and T-distribution. As seen from the figure,  $TCM_\alpha$  is more robust than

$\text{TCE}_\alpha$  in the sense that it is less sensitive to the tail type of the underlying distribution.

### 7.3 Influence Functions of Tail Conditional Expectation and Tail Conditional Median

Influence functions introduced by Hampel [15] [16] is a useful tool in assessing the robustness of an estimator. Let  $F$  be a distribution function. The quantity of interest determined by  $F$  is represented by an estimator,  $T(F)$ , where  $T$  is a functional mapping from a family of distribution functions to  $\mathbb{R}$ . For  $x \in \mathbb{R}$ , let  $\delta_x$  be the point mass 1 at  $x$ . The influence function of the estimator  $T(F)$  at  $x$  is defined by

$$IF(x, T, F) = \lim_{\varepsilon \downarrow 0} \frac{T((1 - \varepsilon)F + \varepsilon\delta_x) - T(F)}{\varepsilon}.$$

The influence function yields information about the rate of change of the estimator  $T(F)$  with respect to a contamination point  $x$  to the distribution  $F$ .

An estimator  $T$  is called bias robust at  $F$ , if its influence function is bounded, i.e.,

$$\gamma^* = \sup_x IF(x, T, F) < \infty.$$

If the influence function of an estimator  $T(F)$  is unbounded, an outlier in the data may cause problems.

**Lemma 3.** Suppose the loss distribution has a density  $f_X(\cdot)$  which is continuous and positive at  $\text{VaR}_{\frac{1+\alpha}{2}}(X)$ , then the influence function of  $\text{TCM}_\alpha$  is given by

$$IF(x, \text{TCM}_\alpha, X) = \begin{cases} \frac{(\alpha - 1)/2}{f_X(\text{VaR}_{\frac{1+\alpha}{2}}(X))}, & x < \text{VaR}_{\frac{1+\alpha}{2}}(X) \\ 0, & x = \text{VaR}_{\frac{1+\alpha}{2}}(X) \\ \frac{(1 + \alpha)/2}{f_X(\text{VaR}_{\frac{1+\alpha}{2}}(X))} & x > \text{VaR}_{\frac{1+\alpha}{2}}(X) \end{cases}$$

Suppose the loss distribution has a density  $f_X(\cdot)$  which is continuous and positive at  $\text{VaR}_\alpha(X)$ , then the influence function of  $\text{TCE}$  is given by

$$IF(x, \text{TCE}_\alpha, X) = \begin{cases} \text{VaR}_\alpha(X) - E[X|X \geq \text{VaR}_\alpha(X)], & \text{if } x \leq \text{VaR}_\alpha(X) \\ \frac{x}{1 - \alpha} - E[X|X \geq \text{VaR}_\alpha(X)] - \frac{\alpha}{1 - \alpha} \text{VaR}_\alpha(X) & \text{if } x > \text{VaR}_\alpha(X) \end{cases} \quad (34)$$

*Proof.* The result of  $\text{TCM}$  is from [30] equation (3.2.3). To show (34), note that by equation (3.2.4) in [30] the influence function of the  $(1 - \alpha)$ -trimmed mean  $T_{1-\alpha}(X) := E[X|X <$

$\text{VaR}_\alpha(X)$  is

$$IF(x, T_{1-\alpha}, X) = \begin{cases} \frac{x - (1-\alpha)\text{VaR}_\alpha(X)}{\alpha} - E[X|X < \text{VaR}_\alpha(X)], & \text{if } x \leq \text{VaR}_\alpha(X) \\ \text{VaR}_\alpha(X) - E[X|X < \text{VaR}_\alpha(X)] & \text{if } x > \text{VaR}_\alpha(X) \end{cases} \quad (35)$$

By simple calculation, the influence function of  $E[X]$  is

$$IF(x, E[X], X) = x - E[X]. \quad (36)$$

Since  $E[X] = \alpha T_{1-\alpha}(X) + (1 - \alpha)\text{TCE}_\alpha$ , it follows that

$$IF(x, E[X], X) = \alpha IF(x, T_{1-\alpha}, X) + (1 - \alpha)IF(x, \text{TCE}_\alpha, X). \quad (37)$$

Then (34) follows from equations (35), (36) and (37).  $\square$

From Lemma 3, we see that

$$\sup_x IF(x, \text{TCM}_\alpha, X) < \infty, \quad \sup_x IF(x, \text{TCE}_\alpha, X) = \infty$$

Hence, TCE has unbounded influence function but TCM has bounded influence function, which implies that TCM is more robust against the influence of outliers in the data.

## 8 Conclusion

We propose a new, data based, risk measure, called natural risk statistic, that are characterized by a new set of axioms. The new axioms only require subadditivity for comonotonic random variables, thus relaxing the subadditivity for all random variables in the coherent risk measure, and the comonotonic additivity in the insurance risk measure. The relaxation is consistent with the prospect theory in psychology.

Comparing to two previous risk measures, the natural risk statistic include tail conditional median which is more robust than tail conditional expectation suggested by the coherent risk measure; and, unlike the insurance risk measure, the natural risk statistic can also incorporate scenario analysis. The natural risk statistic includes the VaR as a special case and therefore show that VaR, though simple, is not irrational.

There are several open problems left. First, the natural risk statistic proposed here is only a static risk measure. It will be of great interest if it can be extended to dynamic risk measures. Furthermore, just like subadditivity in the coherent risk measure has been extended to convex risk measures, we pose a conjecture that the comonotonic subadditivity can be extended to comonotonic convexity.

## References

- [1] Acerbi, C. and C. Nordio and C. Sirtori (2001). Expected shortfall as a tool for financial risk management. Working paper can be downloaded from [http:// www.gloriamundi.org](http://www.gloriamundi.org)
- [2] Acerbi, C. and D. Tasche (2002). On the coherence of Expected Shortfall. *Journal of Banking and Finance*, **26**, 1487-1503.
- [3] Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance*, **9**, 203-228.
- [4] BarndorD-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics (with discussion), *J. Roy. Statist. Soc. Ser. B*, **63**, 167-241.
- [5] Basel Committee on Banking Supervision (1988). *International Covergence of Capital Requirement an Capital Standard*.
- [6] Basel Committee on Banking Supervision (1996). *Amendment to the Capital Accord Incorporate Market Risks*.
- [7] Basel Committee on Banking Supervision (1988). *International Covergence of Capital Requirement an Capital Standard: A Revised Framework*.
- [8] Chicago Mercantile Exchange (1995). Standard Portfolio Analysis of Risk.
- [9] Daniélsson, J., B. N. Jorgensen, G. Samorodnitsky, M. Sarma, C. G. de Vries (2005) Subadditivity Re-examined: the Case for Value-at-Risk. Working paper, available at <http://ideas.repec.org/p/fmg/fmgdps/dp549.html>.
- [10] Denneberg, D. (1994) *Non-additive measure and integral*, Kluwer Academic Publishers, Boston.
- [11] Dhaene, J., R. J. A. Laeven, S. Vanduffel, G. Darkiewicz, and M. J. Goovaerts (2005) Can a Coherent Risk Measure Be Too Subadditive? Working paper, available at <http://www.econ.kuleuven.be/insurance/pdfs/subadditive.pdf>
- [12] Dhaene, J., M. J. Goovaerts, and R. Kaas (2003). Economic Capital Allocation Derived from Risk Measures. *North American Actuarial Journal*. **7**, 44-56.

- [13] Eberlein, E. and U. Keller. (1995). Hyperbolic distributions in finance. *Bernoulli*, **1**, 281-299.
- [14] Embrechts, P., A. Mcneil, and D. Straumann (2001). Correlation and dependency in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, edited by M.A.H. Dempster, Cambridge University Press, U.K.
- [15] Hampel, F. R. (1968) Contributions to the theory of robust estimation. Ph.D. thesis. University of California, Berkeley.
- [16] Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of American Statistical Association*, **62**, 1179-1186.
- [17] Heyde, C. C. and S. G. Kou (2004). On the controversy over tailweight of distributions. *Operations Research Letters*, **32**, 399-408.
- [18] Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.
- [19] Kahneman, D. and A. Tversky (1979) Prospect theory: an analysis of decision under risk. *Econometrica*, **47**, 1979, 263-292.
- [20] Pflug, G. (2000). Some remarks on the Value-at-Risk and the conditional Value-at-Risk. In *Probabilistic Constrained Optimization: Methodology and Applications*, edited by S. Uryasev, Kluwer
- [21] Quiggin, J. (1982) A theory of anticipated utility. *Journal of economic behaviour and organization*, **3**, 323-343.
- [22] Rachev, S. and S. Mittnik (2000). *Stable Paretian Models in Finance*, Wiley, New York.
- [23] Rockafellar, R. T. and S. Uryasev (2002). Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance*, **26**, 1443-1471.
- [24] Wang, S. S., V. R. Young, and H. H. Panjer (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, **21**, 173-183.
- [25] Schmeidler, D. (1986) Integral representation without additivity. *Proceedings of the American Mathematical Society*, **97**, 255-261.

- [26] Schmeidler, D. (1989) Subjective probability and expected utility without additivity, *Econometrica*, **57**, 571-587.
- [27] Tversky, A. and D. Kahneman (1992) Advances in prospect theory: cumulative representation of uncertainty. *Journal of risk and uncertainty*, **5**, 297-323.
- [28] Wakker, P. and A. Tversky (1993) An axiomatization of cumulative prospect theory. *Journal of risk and uncertainty*, **7**, 147-176.
- [29] Yaari, M. E. (1987) The dual theory of choice under risk. *Econometrica*, **55**, 95-115.
- [30] Staudte, R. G., and S. J. Sheather (1990) *Robust Estimation and Testing*, Wiley, New York.