What Is a Good External Risk Measure: Bridging the Gaps between Robustness, Subadditivity, and Insurance Risk Measures

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

June 2006, July 2007

Abstract

Choosing a proper risk measure is of great regulatory importance and is relevant to the interfaces of operations and finance, as exemplified in Basel Accord which uses VaR (or quantiles) in combination with scenario analysis as a preferred risk measure for banking and operational risk. Two main families of axiomatically based risk measures are the coherent risk measures, which assume subadditivity for random variables, and the insurance risk measures, which assume additivity for comonotonic random variables. We propose new, data-based, risk measures, called *natural risk statistics*, that are characterized by a new set of axioms. The new axioms only require subadditivity for comonotonic random variables, consistent with the prospect theory. We point out while many risk measures may be suitable for internal risk management, robustness is an important consideration for external risk measures. Comparing to the previous measures, the natural risk statistics include the tail conditional median which is more robust than the tail conditional expectation suggested by coherent risk measures; and, unlike insurance risk measures, the natural risk statistics can also incorporate scenario analysis. The natural risk statistics includes VaR as a special case and therefore shows that VaR, though simple, is not irrational.

Keywords: Risk measures, utility theory, prospect theory, tail conditional expectation, tail conditional median, value at risk, quantile, robust statistics, L-statistics

1 Introduction

Choosing a proper risk measure is an important regulatory issue relevant to the interfaces of operations and finance, as exemplified in governmental regulations such as Basel Accord [5, 6, 7], which uses VaR (or quantiles) along with scenario analysis as a preferred risk measure for banking and operational risk. The main motivation of the current paper is to investigate whether VaR, in combination with scenario analysis, is suitable to be a good risk measure. By discussing various sets of axioms proposed for risk measures, and by giving a different set of axioms based on data, this paper provides a theoretical basis for using VaR along with scenario analysis as a robust risk measure for the purpose of external, regulatory risk measurement.

1.1 Background

Broadly speaking a risk measure attempts to assign a single numerical value to a 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 regulatory deposit requirements. Consequently, how to choose a good risk measure becomes a problem of great practical importance.

There are two main families of risk measures suggested in the literature, the coherent risk measures suggested by Artzner et al. [3] and the insurance risk measures in Wang et al. [60]. To get a coherent risk measure, one first chooses a set of scenarios (different probability measures), and then computes the coherent risk measure as the maximal expectation of the loss under these scenarios. To get an insurance risk measure, one fixes a distorted probability, and then computes the insurance risk measure as the expectation with respect to one distorted probability (only one scenario). 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 postulated, 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. Here we shall provide a different set of axioms, which is more general and aims at external, regulatory risk measures.

1.2 Objectives of Risk Measures: Internal vs. External Risk Measures

One important issue that have not been well addressed in the existing literature is the objective of choosing a risk measure. More precisely, when we propose a risk measure, is it proposed for the interest of a firm's equity shareholders, regulatory/legal agencies, or the internal management of a firm? There is no reason to believe that there is one unique risk measure fits the needs of these different parties. One risk measure may be suitable for internal management, but not for external regulatory agencies, and vice versa.

In this paper we shall focus on external risk measures from the viewpoint of governmental/regulatory agencies. In particular, to enforce risk measures in governmental regulation, it is desirable to have risk measures that can be implemented consistently throughout all the relevant firms, not matter what internal beliefs or internal models each individual firms may have. More precisely, for external risk measures, we prefer risk measures that are robust with respect to modeling assumptions, and are based on data (could be a mixture of historical data and simulated data generated according to a well-defined procedure agreed by most firms) rather on some subjective internal models. For more background of legal/regulatory requirements, see Sections 3 and 4.

1.3 Contribution of This Paper

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 sevenfold.

(1) We give reasons on why a different set of axioms is needed: (a) We point out some critiques of subadditivity mainly from robustness view point (see Section 4), as well as from utility theory and psychological viewpoints (see Section 5). (b) The main drawback of insurance risk measure is that it does not incorporate scenario analysis; i.e. unlike the coherent risk measures, an 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 (according to some generally agreed procedure) or both, rather than on some hypothetical distributions.

(2) A different set of axioms based on data and comonotonic subadditivity is postulated in Section 6, resulting in the definition of *natural risk statistics*. A complete characterization of the natural risk statistics is given in Theorem 1.

(3) An alternative characterization of the natural risk statistics based on statistical acceptance sets is given in Theorem 2 in Section 6.2.

(4) VaR or quantiles in combination of scenario analysis is among the most widely used risk measures in practice (see, e.g. the Basel Accord). However the coherent risk measures rule out the use of VaR. In Section 7 we show that the natural risk statistics give an axiomatic justification to the use of VaR in combination of scenario analysis.

(5) Theorems 3 and 4 in Section 7.1 completely characterize data-based coherent risk measures and data-based law-invariant coherent risk measure. As suggested in Theorem 4,

natural risk statistics are in general more robust than coherent risk measures.

(6) Theorem 5 in Section 7.2 completely characterizes data-based insurance risk measures. Unlike the insurance risk measures, the natural risk statistics can incorporate scenario analysis by putting different set of weights on the sample order statistics.

(7) We point out in Section 8 that the natural risk statistics include the *tail conditional median* as a special case, which leads to a more robust measure of risk than the tail conditional mean suggested by the coherent risk measures.

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 [33] 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. In addition, we need to show that the weights are nonnegative and add up to one.

2 Review of Existing Risk Measures

2.1 Coherent and Convex Risk Measures

Let Ω be the set of all possible states at the end of an observation period, and $\mathcal{X}$ be the set of financial losses under consideration. Then a risk measure ρ is a mapping from $\mathcal{X}$ to $\mathbb{R}$.

2.1.1 Subadditivity

Artzner et al. [3] proposed risk measures based on subadditivity. In particular, Artzner et al. [3] called a risk measure ρ 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. There are at least two different types of risk measures, accounting-based risk measures and attitude-based (related to utility functions) risk measures. Axiom A1, which says that a sure loss of amount b simply increases the risk by b, is mainly an axiom for accounting-based risk measures. For many external risk measurs, such as margin deposit, the accounting-based risk measures seem to be reasonable. For internal risk measures, attitude-based risk measures may be prefered. To get attitude-based risk meausre, one replace Axiom A1 by other axioms, such as the ones in the convex risk measures[1], in which case a sure loss of amount b does not 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" (Artzner et al. [3], p. 209). We will discuss the controversies related to this axiom in Section 5.

Artzner et al. [3] pointed out that Huber [33] showed that if Ω has a finite number of elements and $\mathcal{X}$ is the set of all real random variables, then a risk measure ρ is coherent if and only if there exists a family $\mathcal{Q}$ of probability measures on Ω, such that

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

where $E^Q[X]$ is the expectation of X under the probability measure Q. Delbaen [15] extended the above result when Ω has infinite number of elements[2]. Therefore, getting a coherent risk measure amounts to computing maximal expectation under different scenarios (different Q's), thus justifying scenarios analysis used in practice. Artzner et al. [3] and Delbaen [15] also presented an equivalent approach of defining the coherent risk measure through acceptance sets.

2.1.2 Law Invariance

A desirable property of risk measures is called law invariance, as stated in the following axiom:

Axiom A4. Law invariance: $\rho(X) = \rho(Y)$, if X and Y have the same distribution under probability measure P.

[1] Convex risk measures were proposed by Föllmer and Schied [23] and independently by Frittelli and Gianin [26] where the positive homogeneity and subadditivity axioms are relaxed to a single convexity axiom: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$, for any $X, Y \in \mathcal{X}$, $\lambda \in [0,1]$. Law-invariant convex risk measures are discussed further in Föllmer and Schied [24], Dana [14], Frittelli and Gianin [27], Ruschedorf [48] and Schied [50].

[2] Consider a random loss X defined on $\mathcal{X} = L^\infty(\Omega, \mathcal{F}, P)$ with $(\Omega, \mathcal{F}, P)$ being a general probability space. A risk measure $\rho : L^\infty(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is said to satisfy the Fatou property if $\rho(X) \leq \liminf_{n\to\infty} \rho(X_n)$, for any sequence $(X_n)_{n\geq 1}$ uniformly bounded by 1 and converging to X in probability. Delbaen [15] showed that $\rho : L^\infty(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is a coherent risk measure with the Fatou property if and only if there exists a closed and convex set of probability measures $\mathcal{Q}$, all of them being absolutely continuous with respect to P, such that $\rho(X) = \sup_{Q\in\mathcal{Q}}\{E^Q[X]\}$, $\forall X \in L^\infty(\Omega, \mathcal{F}, P)$.

Law invariance means that the risk of a position is determined purely by the loss distribution. A risk measure is called a law-invariant coherent risk measure, if it satisfies Axiom A1-A4. Kusuoka [40] gives a representation for the law-invariant coherent risk measures[3].

2.1.3 Tail Conditional Expectation

A special case of coherent risk measures, the tail conditional expectation (TCE) 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 [2]. The TCE is also called expected shortfall by Acerbi et al.[1] and conditional value-at-risk by Rockafellar and Uryasev [46] and Pflug [42]. More precisely, the TCE at level α is defined by

$$\text{TCE}_{\alpha}(X) = \text{mean of the } \alpha\text{-tail distribution of } X. \tag{1}$$

If the distribution of X is continuous, then

$$\text{TCE}_{\alpha}(X) = E[X|X \geq \text{VaR}_{\alpha}(X)]. \tag{2}$$

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

2.1.4 Main Drawbacks

A main drawback of the coherent risk measures and convex risk measures is that it includes TCE as a risk measure. This is fine for internal risk measures. However, as we shall point out in Sections 4 and 5, using TCE is troublesome for external risk measures, mainly because TCE is too sensitive to the modeling assumptions for tail distributions. Thus, it is very difficult to implement TCE consistently as part of governmental regulations.

Another drawback of coherent risk measures and convex risk measures is that they rule out the use of quantiles as risk measures, as they are not expectations. One of the most widely used risk measures in regulations of risk management is Value-at-Risk (VaR), which is nothing but a quantile at some pre-defined probability level. More precisely, given

[3] Let F_X be the distribution function for a random variable $X \in L^{\infty}(\Omega, \mathcal{F}, P)$. Define $F_X^{-1}(u) = \inf\{x : F_X(x) > u\}$, $u \in [0,1)$, $V_{\alpha}(X) = \frac{1}{\alpha}\int_{1-\alpha}^{1} F_X^{-1}(u)du$, $X \in L^{\infty}(\Omega, \mathcal{F}, P)$, $\alpha \in (0,1]$. Kusuoka [40] proved that if P is a non-atomic probability measure, then $\rho : L^{\infty}(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is a law-invariant coherent risk measure with the Fatou property if and only if there exists a compact convex set $\mathcal{M}$ of probability measures on $[0,1]$, such that $\rho(X) = \sup_{m \in \mathcal{M}}\{\int_0^1 V_{\alpha}(X)m(d\alpha)\}$, $\forall X \in L^{\infty}(\Omega, \mathcal{F}, P)$.

$\alpha \in (0, 1)$, the value-at-risk VaR_α at level α of the loss variable X is defined as the α-quantile of X, i.e.,

$$\mathrm{VaR}_\alpha(X) = \min\{x \mid P(X \le x) \ge \alpha\}. \tag{3}$$

For example the banking regulation "Basel Accord" specifies a preferred risk measure as VaR at 99 percentile under various scenarios.

Therefore, the very fact that coherent risk measures and convex risk measures exclude VaR and quantiles posts a serious inconsistency between the academic theory and governmental practice. 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 5.

By relaxing this axiom and requiring subadditivity only for comonotonic random variables, we are able to find a new set of axioms in Section 6 which will include VaR and quantiles, thus eliminating this inconsistency.

2.2 Insurance Risk Measures

Insurance risk premiums can also be viewed as risk measures, as they aim at using one numerical number to summarize future random losses. To characterize insurance risk premiums, Wang et al. [60] proposed four axioms; more precisely, a risk measure ρ is said to be an insurance risk measure if it satisfies the following five axioms.

Axiom B1. Law invariance: the same as Axiom A4.

Axiom B2. Monotonicity: $\rho(X) \le \rho(Y)$, if $X \le Y$ almost surely.

Axiom B3. Comonotonic additivity: $\rho(X + Y) = \rho(X) + \rho(Y)$, if X and Y are comonotonic, where random variables X and Y are comonotonic if and only if

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

holds almost surely for ω_1 and ω_2 in Ω.

Axiom B4. Continuity:

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

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

Axiom B5. Scale normalization: $\rho(1) = 1$.

The notion of comonotonic random variables is discussed in Schmeidler [51], Yaari [62] and Denneberg [16]. The psychological motivation of comonotonic random variables comes from Prospect Theory; see Section 5.3. See Dhaene et. al. [19] for a recent review of risk measures and comonotonicity.

If two random variables X and Y are comonotonic, then $X(\omega)$ and $Y(\omega)$ always move in the same direction as the state ω changes. For example, the payoffs of a call option and its underlying asset are comonotonic.

Wang et al. [60] imposed Axiom B3 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 7 will show that comonotonic additivity fails to hold.

Wang et al. [60] proved that if $\mathcal{X}$ contains all the $Bernoulli(p)$ random variables, $0 \leq p \leq 1$, then risk measure ρ satisfies axioms B1-B5 if and only if ρ has a Choquet integral representation with respect to a distorted probability:

$$\rho(X) = \int X d(g \circ P) = \int_{-\infty}^{0} (g(P(X > t)) - 1)dt + \int_{0}^{\infty} g(P(X > t))dt, \tag{4}$$

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 [16].

It should be emphasized that VaR satisfies the axioms B1-B5 (see Corollary 4.6 in Denneberg [16] for a proof that VaR satisfies Axiom B4) 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, an insurance risk measure in (4) does not satisfy subadditivity, unless the distortion function $g(\cdot)$ is concave (see Denneberg [16]).

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

The main reason that insurance risk measures rule out scenario analysis is that they require comonotonic additivity. The counterexample at the end of Section 7 shows that even for comonotonic random variables, with different scenarios we may get strict subadditivity rather than additivity. In our new approach in Section 6 we shall require comonotonic subadditivity instead of comonotonic additivity.

The mathematical concept of comonotonic subadditivity was also studied independently by Song and Yan [53], who gave a representation of the functionals satisfying comonotonic subadditivity or comonotonic convexity from a mathematical perspective[4]. In [54], they gave a representation of risk measures that are not only comonotonically subadditive or convex, but also respect stochastic orders.

There are several differences between Song and Yan [53],[54] and the current paper. First, our paper provides a full mathematical characterization of the new risk statistics, which are based on data (either observed or simulated or both) rather than on some hypothetical distributions. Second, we provide economic, psychological, and legal reasons for postulating the comonotonic subadditivity axiom, not just for mathematical convenience. Third, we give two representations of the data-based coherent and insurance risk measures in Section 7, so that we can compare the new risk measures with existing risk measures. Fourth, we provide alternative axioms for risk measures based on acceptance sets.

2.3 Static vs. Dynamic Risk Measures

It should be emphasized that, similar to coherent and insurance risk measures, in this paper we only consider static risk measures, i.e., one period risk measures. Readers interested in dynamic risk measures, which have more controversies, are referred to, e.g., [4, 44, 41, 47, 61, 17, 11, 8, 37, 25, 12].See in particular, some counterexamples given in [9].

3 Philosophical Basis of Our Arguments and Basic Concepts of the Law

In this section, we shall summarize some basic concepts of the law, which motivate us to propose a new set of axioms for risk measures used for external regulation. By definition, an axiom is "a statement or principle which is generally accepted to be true, but is not necessarily so" (Cambridge English Dictionary). Hence, axioms are subject to debate and change. Alternative axioms are therefore useful because they provide people with more choices and they may be more suitable than existing axioms in certain circumstances.

Just like there are key differences between *internal standards* (such as morality) and *external standards* (such as law and regulation), there are differences between internal and

[4] They proved that a functional ρ defined on $L^\infty(\Omega, \mathcal{F})$ or $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ which satisfies Axiom A1, A2 and comonotonic subadditivity can be represented by $\rho(X) = \max_{\mu \in \mathcal{M}} \mu(X)$, where $\mathcal{M}$ is a certain set of monotonic set functionals on $\mathcal{F}$ and $\mu(X)$ is the Choquet integral of X with respect to μ. Similar representations are proved for comonotonic convex functionals.

external risk measures. By understanding basic concepts of the law, we will have a better understanding of what are needed in axioms proposed for external risk measures. We shall point out: (1) Risk measures used in external regulation should be robust because robustness is essential for law enforcement. (2) Risk measures used in external regulation should have consistency with people's behavior because law should reflect society norms.

There is a vast literature on the philosophy of law (see, e.g., Hart [29]). Two concepts to be discussed here are *legal realism* and *legal positivism*. The former requests the robustness of the law, and the latter requests the consistency of the law to social norms.

3.1 Legal Realism and Robustness of the Law

Legal realism is the viewpoint that the legal decision of a court regrading to a case is determined by the actual practices of judges, rather than the law set forth in statutes or precedents. All legal rules contained in both statutes and precedents have uncertainty due to the uncertainty in human language and inability to anticipate all possible future circumstances([29], p. 128). Hence a law is only a guideline for judges and enforcement officers ([29], p. 204-205), i.e., a law is only intended to be the average of what the judges and officers will decide. This requests the *robustness* of the law, i.e., we hope that different judges will arrive at similar conclusions when they implement the law.

In particular, the enforcement of risk measures in banking regulation requires that risk measures should be robust with respect to the underlying models and data. However, the coherent risk measures generally lack robustness, as discussed in Section 4 and manifested in Theorem 4 of Section 7.1. The further study in Section 8 shows that the tail condition median, a special case of our proposed new risk measures, is more robust than the tail conditional expectation.

3.2 Legal Positivism and Social Norm

Legal positivism is the thesis that the existence and content of law depend on social norms and not on their merits, mainly because if a system of rules are to be imposed by force in the form of law, there must be a sufficient number of people who accept it voluntarily. Without their voluntary cooperation, the coercive power of law and government cannot be established ([29], p. 201-204).

Therefore, risk measures imposed in banking regulations should also reflect most people's behavior; otherwise, the regulation cannot be enforced. However, the study of the prospect theory in psychology showed that in face of financial risk, most people's decision can violate the subadditivity Axiom A3 (see Section 5.3 for details). This motivated us to propose the new risk measure, which is consistent with most people's behavior.

3.3 An Example of Speed Limit

An illuminating example manifesting the above ideas is the setting up of speed limit on the road, which is a crucial issue involving life and death decisions. In 1974, the U.S. Congress enacted a National Maximum Speed Law that federally mandated that no speed limit may be higher than 55 mph. The law was widely disregarded, even after the national maximum was increased to 65 mph in 1987 on certain roads. In 1995, the law was repealed, returning the choice of speed limit to each state, in part because of notoriously low compliance.

Today, the "Manual on Uniform Traffic Control Devices" of AASHTO (American Association of State Highway and Transportation Officials) recommends setting speed limit near the 85th percentile speed of free flowing traffic (see [57], p. 51) with an adjustment taking into consideration that people tend to drive 5 to 10 miles above the posted speed limit. This recommendation is adopted by all states and most local agencies [35]. Although the 85th percentile rule appears to be a simple method, studies have shown that crash rates are lowest at around the 85th percentile.

The 85th percentile speed manifests the robustness of law and its consistency to social norms: (1) The 85th percentile rule is robust in the sense that it is based on data rather than on some subjective models, and it can be implemented consistently. (2) Laws that reflect the behavior of the majority of drivers are found to be successful, while laws that lack the consent and voluntary compliance of the public majority cannot be effectively enforced.

4 The Main Reason to Relax Subadditivity: Robustness

When a regulator imposes a risk measure, it must be unambiguous, stable, and can be implemented consistently throughout all the relevant firms. Otherwise, different firms using different models may report very different risk measures to the regulator; even worse, some firms may even game the system in relatively easy ways. In short, from a regulator viewpoint, the risk measure should demonstrate robustness with respect to underlying models, in order to enforce the regulation and to maintain the stability of the regulation.

The robustness of coherent risk measures based on subadditivity is questionable:

(1) The theory of coherent risk measures suggests to use the tail conditional expectation (TCE) to compute risk measures. However, the TCE may be sensitive to model assumptions of heaviness of tail distributions, which is a controversial subject.

For example, although it is accepted that stock returns have tails heavier than those of normal distribution, one school of thought believes tails to be exponential type and another believes power-type tails. Heyde and Kou [32] 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. If the percentiles have to be estimated from data, then the situation is even worse, as we have to rely on confidence intervals which may have significant overlaps. Therefore, with ordinary sample sizes (e.g. 20 years of daily data), one cannot easily identify exact tail behavior from data.

In summary, the tail behavior may be a subjective issue depending on people's modeling preferences. Since as we will show in Section 8 that TCE is sensitive to the assumption on the tail distribution behavior, using TCE as an external risk measure can be problematic if the tail behavior is a subjective issue.

(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 $\tilde{x} = (x_1, \ldots, x_n)$ from a random loss X, let $(x_{(1)}, \ldots, x_{(n)})$ 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 in data and to model assumptions in simulation and analysis.

More generally, let $\tilde{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n$ be a weight with $0 \leq w_1 \leq w_2 \leq \cdots \leq w_n$ and $\sum_{i=1}^n w_i = 1$. Then the risk measure $\hat{\rho}(\tilde{x}) = \sum_{i=1}^n w_i x_{(i)}$ is an empirically coherent risk measure satisfying subadditivity, as will be shown in Section 7. However, since this risk measure puts larger weights on larger observations, it is obviously not robust. In fact, as we will prove in Theorem 4 in Section 7, any empirically law-invariant coherent risk measure $\hat{\rho}(\tilde{x})$ can be represented by

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^n w_i x_{(i)}\},$$

where $\mathcal{W} = \{\tilde{w}\}$ is a set of weights with each $\tilde{w} = (w_1, \ldots, w_n) \in \mathcal{W}$ satisfying $w_1 \leq$

$w_2 \leq \cdots \leq w_n$, which is not robust as it puts more weights on more extreme observations. Therefore, coherent risk measures are generally not robust. We will discuss the issue of robustness in more detail in Section 8.

5 Other Reasons to Relax Subadditivity

5.1 Diversification And Tail Subadditivity of VaR

Both the subadditivity axiom and the convexity axiom conform to the idea that diversification does not increase the risk. There are two main motivations for diversification. One is based on the simple observation that $SD(X+Y) \leq SD(X) + SD(Y)$, for any two random variables X and Y with finite second moments, where $SD(\cdot)$ denotes standard deviation. The other is based on expected utility theory. Samuelson [49] showed that any investor with a strictly concave utility function will uniformly diversify among i.i.d. risks with finite second moments, i.e., the expected utility of the uniformly diversified portfolio is larger than that of any other portfolio. Both of the two motivations require finiteness of second moments of the risks.

Is diversification still preferable for risks with infinite second moments? The answer can be no. Ibragimov and Walden [34] showed that diversification is not preferable for unbounded extremely heavy-tailed distributions, in the sense that the expected utility of the diversified portfolio is smaller than that of the undiversified portfolio[5]. Ibragimov and Walden [34] also showed that, investors with certain S-shaped utility functions would prefer non-diversification, even for bounded risks.[6]. A S-shaped utility function is convex in the domain of loss. The convexity in the domain of loss is supported by experimental results and the prospect theory [38] [58], which is an important alternative to the expected utility theory. We will have more discussion about prospect theory in section 5.3.

The fact that diversification is not universally preferable makes it unreasonable to criticize VaR just because it does not have subadditivity universally. Although in the center

[5] Let $X_i, i = 1, \ldots, n$ be i.i.d. risks with unbounded heavy-tail distribution belonging to the class $\mathcal{CS}(r)$(see [34] for definition) with $r < 1$, in particular $E|X_i| = \infty$. Let $X_w \triangleq \sum_{i=1}^n w_i X_i$ be the diversified portfolio of the risks, where $w \in \mathbb{R}^n$ is a nonnegative weight with $\sum_{i=1}^n w_i = 1$. Let X^a denote the truncation of random variable X on $[-a, a]$, i.e., $X^a \triangleq \max\{-a, \min\{X, a\}\}$, where $a > 0$. They showed that there is an a_0 such that for all $a > a_0$ and all concave utility function u, it holds that $E(u(X_1^a)) \geq E(u(X_w^a))$, i.e., investor would prefer single risk X_1 instead of diversified risk X_w.

[6] Let $X_i, i = 1, \ldots, n$ be i.i.d. risks with unbounded heavy-tail distribution belonging to class $\mathcal{CS}(r)$. Let X_i^a be the truncation of X_i. Then X_i^a are bounded i.i.d. risks. Let $\bar{X}(a) = \frac{1}{n}\sum_{i=1}^n X_i^a$ be the diversified portfolio. They showed that there exist S-shaped utility functions v such that $E(v(X_1^a)) > E(v(\bar{X}(a)))$.

of the distributions VaR may violate the subadditivity, Daníelsson et al. [13] questioned whether the violation is merely a technical issue, at least if one focuses on the tail regions which are the most relevant regions for risk management. Indeed they showed that VaR is subadditive in the tail regions, provided that the tails in the joint distribution are not extremely fat (with tail index less than one)[7]. They also carried out simulations showing that VaR_α is indeed subadditive when $\alpha \in [95\%, 99\%]$ for most practical applications.

To summarize, there is no conflict between the use of VaR and diversification. When the risks do not have extremely heavy tails, diversification is preferred and VaR satisfies subadditivity in the tail region; when the risks have extremely heavy tails, diversification may not be preferable and VaR may fail to have subadditivity.

	Not Very Fat Tails	Fat Tails
Does diversification help to reduce risk?	Yes	No
Does VaR satisfy subadditivity?	Yes	No

Asset returns with tail index less than one have very fat tails. They are hard to find and easy to identify. Daníelsson et al. [13] argued that they can be treated as special cases in financial modeling. Even if one encounters an extreme fat tail and insists on tail subadditivity, Garcia et al. [28] showed that, when tail thickness causes violation of subadditivity, a decentralized risk management team may restore the subadditivity for VaR by using proper conditional information.

5.2 Does Merger Always Reduce Risk

Subadditivity basically means that "a merger does not create extra risk" (Artzner et al. [3], p. 209). However, Dhaene et al. [18] pointed out that many times merger may increase risk, particularly due to bankruptcy protections for 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 risk and merger may increase risk[8].

[7] More precisely, Daníelsson et al. [13] 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 $\limsup_{\alpha \to 1} \frac{\text{VaR}_\alpha(X+Y)}{\text{VaR}_\alpha(X)+\text{VaR}_\alpha(Y)} \leq 1$.

[8] Mathematically, let X and Y be the net payoff of two firms before a merger. Because of the bankruptcy protection, the actual net payoff of the two firms would be $X^+ = \max(X, 0)$ and $Y^+ = \max(Y, 0)$, respectively. After the merger, the net payoff of the joint firm would be $X + Y$, and the actual net payoff would

For example, 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 sunk the entire bank.

In addition, there is little empirical evidence supporting the argument that "a merger does not create extra risk". Indeed, in practice, the credit agencies, such as Moody's and Standard & Poor's, will not upgrade a firm's credit rating because of a merger; on the contrary, many times the credit rating of the joint firm may be lower shortly after the merger of two firms.

5.3 Reasons 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. Kahneman and Tversky [38] proposed a model of choice under uncertainty called "prospect theory," leading to a Nobel prize in Economics. In particular, many people have studied on using comonotonic random variables; see Kahneman and Tversky [58], Tversky and Wakker [59], Quiggin [43], Schmeidler [51], [52] and Yaari [62]. These 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.

There are simple examples showing that risk associated with non-comonotonic random variables may violate subadditivity, because people are risk seeking for the losses of moderate or high probability, as implied by prospect theory[9].

be $(X+Y)^+$, due to bankruptcy protection. Because $(X+Y)^+ \leq X^+ + Y^+$, a merger always results in a decrease in the actual net payoff, if one only considers the effect of bankruptcy protection in a merger. In other words, a merger increases the risk of investment given everything else being equal. This contradicts the intuition that "A merger does not create extra risk".

[9] Suppose there is an urn which contains 50 black balls and 50 red balls. Let B be the event of losing \$10,000 if a ball randomly drawn from the urn is black, and R be the event of losing \$10,000 if a ball randomly drawn from the urn is red. Obviously, B and R have the same measure of risk, i.e., $\rho(B) = \rho(R)$. Let S be the event of losing \$5,000 for sure, then $\rho(S) = 5,000$. According to the prospect theory, people are risk seeking for the losses of moderate or high probability, i.e., most people would prefer a substantial

Schmeidler [52] 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 $\rho(X+Y) \le \rho(X) + \rho(Y)$ only for comonotonic random variables X and Y.

The insurance risk measures impose comonotonic additivity in Axiom B3, 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 7 shows that if one incorporates different scenarios, then additivity may not hold even for comonotonic random variables. Hence, the comonotonic additivity condition in Axiom B3 may be too restrictive and its relaxation to comonotonic subadditivity may be a better choice.

5.4 Superadditivity vs. Subadditivity

In terms of utility theory, it is not clear whether a risk measure should be superadditive or subadditive, at least for independent random variables. For example, Hennessy and Lapan [31] show that for utilities with increasing relative risk aversion, two individual lotteries may be perferable to the summed lottery, i.e., risk measures of lotteries can be superadditive[10]. In an interesting paper Eeckhoudt and Schlesinger [21] link the sign of utility function to risk preferences[11]. Since we shall impose subadditivity only for commonotonic random

probability of a larger loss over a sure loss. Therefore, most people would prefer position B over position S (see problem 12 on p. 273 in Kahneman and Tversky [38], and table 3 on p. 307 in Tversky and Kahneman [58]). In other words, we have $\rho(B) = \rho(R) < \rho(S) = 5,000$. On the other hand, since the position $B+R$ corresponds to a sure loss of \$10,000, we have $\rho(B+R) = 10,000$. Combining together we have $\rho(B)+\rho(R) < 5,000+5,000 = 10,000 = \rho(B+R)$, violating the subadditivity. Clearly the random losses associated with B and R are not comonotonic. Therefore, this example shows that risk associated with non-comonotonic random variables may not have subadditivity. Schmeidler [52] attributes this phenomena 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.

[10] Let $u(\cdot)$ be the utility function and $ce(\cdot)$ be the functional of certainty equivalent, i.e., $ce(X) \triangleq u^{-1}(E(u(X)))$, for any lottery X. Hennessy and Lapan [31] proved that if u has increasing relative risk aversion, then there exist $X \ge 0$ and $Y \ge 0$, such that $ce(X+Y) < ce(X) + ce(Y)$. In other words, $X+Y$ has less utility and hence larger risk.

[11] Let $u^{(4)}$ be the fourth derivative of an utility function u. They proved that $u^{(4)} \le 0$ if and only if $E(u(x+\epsilon_1+\epsilon_2)) + E(u(x)) \le E(u(x+\epsilon_1)) + E(u(x+\epsilon_2))$, for any $x \in \mathbb{R}$ and any independent risk ϵ_1 and ϵ_2 such that $E(\epsilon_1) = E(\epsilon_2) = 0$. This result can be interpreted as follows. Suppose the owner of two sub-firms, each of which has initial wealth x, faces the problem of assigning two projects to the sub-firms. The net payoff of the two projects are ϵ_1 and ϵ_2, respectively. The result suggests that, whether the owner prefers to assign both projects to a single subfirm or prefers to assign one project to each sub-firm depends on the sign of the fourth derivative of his utility function.

variables, there is no contradiction with the above mentioned results in utility theory because commontonic random variables are not independent random variables.

6 Main Result: Natural Risk Statistics and Their Axiomatic Representations

6.1 The First Representation

In this section we shall propose a new measure of risk based on data. Suppose we have a collection of data observation $\tilde{x} = (x_1, x_2, ...x_n) \in \mathbb{R}^n$ 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 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 $\mathbb{R}$. 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 simulated or both) $\tilde{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ from X. Next we shall introduce a set of axioms for $\hat{\rho}$.

Axiom C1. Positive homogeneity and translation invariance:

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

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

Axiom C2. Monotonicity: $\hat{\rho}(\tilde{x}) \leq \hat{\rho}(\tilde{y})$, if $\tilde{x} \leq \tilde{y}$, where $\tilde{x} \leq \tilde{y}$ means $x_i \leq y_i, i = 1, \ldots, n$.

The above two axioms have been proposed for the coherent risk measures. Here we simply adapted them to the case of risk statistics. Note that Axiom C1 yields

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

Note that we can easily relax the requirement of $\hat{\rho}(b\mathbf{1}) = b$. For example, if we require the loss suggested from a risk measure cannot exceed 10% of the total capital, then we can simply require $\hat{\rho}(b\mathbf{1}) = b/0.1 = 10 * b$. Also axioms C1 and C2 imply $\hat{\rho}$ is continuous[12].

Axiom C3. Comonotonic subadditivity:

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

[12] Indeed, suppose $\hat{\rho}$ satisfies axioms C1 and C2. Then for any $\tilde{x} \in \mathbb{R}^n$, $\varepsilon > 0$, and $\tilde{y}$ satisfying $|y_i - x_i| < \varepsilon, i = 1, \ldots, n$, we have $\tilde{x} - \varepsilon\mathbf{1} < \tilde{y} < \tilde{x} + \varepsilon\mathbf{1}$. By the monotonicity in Axiom C2, we have $\hat{\rho}(\tilde{x} - \varepsilon\mathbf{1}) \leq \hat{\rho}(\tilde{y}) \leq \hat{\rho}(\tilde{x} + \varepsilon\mathbf{1})$. Applying Axiom C1, 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 $\hat{\rho}$.

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

Axiom C4. Permutation invariance:

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

This axiom can be considered as the counterpart of the law invariance Axiom A4 in terms of data. It means that if two data $\tilde{x}$ and $\tilde{y}$ have the same empirical distribution, i.e., the same order statistics, then $\tilde{x}$ and $\tilde{y}$ should give the same estimate of risk. It is postulated because we focus on risk measures of a single random variable X with data observation $\tilde{x}$. In other words, just like coherent risk measures and insurance risk measures, here we discuss static risk measures rather than dynamic risk measures.

Definition 1. A risk statistic $\hat{\rho} : \mathbb{R}^n \to \mathbb{R}$ is called a natural risk statistic if it satisfies axioms C1-C4.

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

Theorem 1. Let $x_{(1)}, ..., x_{(n)}$ be the order statistics of the observation $\tilde{x}$ with $x_{(n)}$ being the largest.

(I) For an arbitrarily given set of weights $\mathcal{W} = \{\tilde{w} = (w_1, \ldots, w_n)\} \subset \mathbb{R}^n$ with each $\tilde{w} \in \mathcal{W}$ satisfying $\sum_{i=1}^n w_i = 1$ and $w_i \geq 0$ for $i = 1, \ldots, n$, the risk statistic

$$\hat{\rho}(\tilde{x}) \triangleq \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^n w_i x_{(i)}\}, \ \forall \tilde{x} \in \mathbb{R}^n \tag{5}$$

is a natural risk statistic.

(II) If $\hat{\rho}$ is a natural risk statistic, then there exists a set of weights $\mathcal{W} = \{\tilde{w} = (w_1, \ldots, w_n)\} \subset \mathbb{R}^n$ with each $\tilde{w} \in \mathcal{W}$ satisfying $\sum_{i=1}^n w_i = 1$ and $w_i \geq 0$ for $i = 1, \ldots, n$, such that

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^n w_i x_{(i)}\}, \ \forall \tilde{x} \in \mathbb{R}^n. \tag{6}$$

Proof. See the on-line supplement. □

The main difficulty in proving Theorem 1 lies in part (II). Axiom C3 implies that the functional $\hat{\rho}$ satisfies subadditivity on comonotonic sets of $\mathbb{R}^n$, for example, on the set $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \le y_2 \le \cdots \le y_n\}$. However, unlike in the case of coherent risk measures, the existence of a set of weights $\mathcal{W}$ such that (6) holds does not follow easily from the proof in Huber [33]. 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 very cautious when using the results of separating hyperplanes. Furthermore, we have to spend some effort showing that $w_i \ge 0$ for $i = 1, \ldots, n$.

6.2 The Second Representation via Acceptance Sets

An alternative view of risk is to define risk as something that may not be acceptable. Similar to coherent risk measures, we shall show the proposed natural risk statistics can also be characterized via acceptance sets. More precisely, 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

$$\hat{\rho}_{\mathcal{A}}(\tilde{x}) = \inf\{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{A}\}, \ \forall \tilde{x} \in \mathbb{R}^n. \tag{7}$$

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

$$\mathcal{A}_{\hat{\rho}} = \{\tilde{x} \in \mathbb{R}^n \mid \hat{\rho}(\tilde{x}) \le 0\}. \tag{8}$$

Thus, one can go from a risk measure to an acceptance set, and vice versa.

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

Axiom D1. The acceptance set $\mathcal{A}$ contains $\mathbb{R}^n_-$ where $\mathbb{R}^n_- = \{\tilde{x} \in \mathbb{R}^n \mid x_i \le 0, i = 1, \ldots, n\}$.

Axiom D2. The acceptance set $\mathcal{A}$ does not intersect the set $\mathbb{R}^n_{++}$ where $\mathbb{R}^n_{++} = \{\tilde{x} \in \mathbb{R}^n \mid x_i > 0, i = 1, \ldots, n\}$.

Axiom D3. 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 D4. 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 $\lambda \ge 0$.

Axiom D5. If $\tilde{x} \le \tilde{y}$ and $\tilde{y} \in \mathcal{A}$, then $\tilde{x} \in \mathcal{A}$.

Axiom D6. If $\tilde{x} \in \mathcal{A}$, then $(x_{i_1}, \ldots, x_{i_n}) \in \mathcal{A}$ for any permutation $(i_1, \ldots, i_n)$.

We will show that a natural risk statistic and a statistical acceptance set satisfying axioms D1-D6 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 D1-D6.

(II) If a statistical acceptance set $\mathcal{A}$ satisfies axioms D1-D6, 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 D1-D6, then $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}} = \bar{\mathcal{D}}$, the closure of $\mathcal{D}$.

Proof. See the on-line supplement. □

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 coherent risk measures in Artzner et al. [3].

7 Comparison between Natural Risk Statistics, Coherent Risk Measures and Insurance Risk Measures

7.1 Comparison with Coherent Risk Measures

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

Definition 2. A risk statistic $\hat{\rho} : \mathbb{R}^n \to \mathbb{R}$ is called a coherent risk statistic, if it satisfies axioms C1, C2 and the following Axiom E3:

Axiom E3. Subadditivity: $\hat{\rho}(\tilde{x} + \tilde{y}) \le \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y})$, for every $\tilde{x}, \tilde{y} \in \mathbb{R}^n$.

We have the following representation theorem for coherent risk statistics.

Theorem 3. A risk statistic is a coherent risk statistic if and only if there exists a set of weights $\mathcal{W} = \{\tilde{w} = (w_1, \ldots, w_n)\} \subset \mathbb{R}^n$ with each $\tilde{w} \in \mathcal{W}$ satisfying $\sum_{i=1}^{n} w_i = 1$ and $w_i \ge$

$0, i = 1, \ldots, n$, such that

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^{n} w_i x_i\}, \ \forall \tilde{x} \in \mathbb{R}^n. \tag{9}$$

Proof. See the on-line supplement. □

Natural risk statistics require the permutation invariance, which is not required by coherent risk statistics. To have a complete comparison between natural risk statistics and coherent risk measures, we consider the following law-invariant coherent risk statistics, which is the counterpart of law-invariant coherent risk measures in the literature.

Definition 3. A risk statistic $\hat{\rho} : \mathbb{R}^n \to \mathbb{R}$ is called a law-invariant coherent risk statistic, if it satisfies axioms C1, C2, C4 and E3.

We have the following representation theorem for the law-invariant coherent risk statistics.

Theorem 4. Let $x_{(1)}, ..., x_{(n)}$ be the order statistics of the observation $\tilde{x}$ with $x_{(n)}$ being the largest.

(I) For an arbitrarily given set of weights $\mathcal{W} = \{\tilde{w} = (w_1, \ldots, w_n)\} \subset \mathbb{R}^n$ with each $\tilde{w} \in \mathcal{W}$ satisfying

$$\sum_{i=1}^{n} w_i = 1, \tag{10}$$

$$w_i \geq 0, i = 1, \ldots, n, \tag{11}$$

$$w_1 \leq w_2 \leq \ldots \leq w_n, \tag{12}$$

the risk statistic

$$\hat{\rho}(\tilde{x}) \triangleq \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^{n} w_i x_{(i)}\}, \ \forall \tilde{x} \in \mathbb{R}^n \tag{13}$$

is a law-invariant coherent risk statistic.

(II) If $\hat{\rho}$ is a law-invariant coherent risk statistic, then there exists a set of weights $\mathcal{W} = \{\tilde{w} = (w_1, \ldots, w_n)\} \subset \mathbb{R}^n$ with each $\tilde{w} \in \mathcal{W}$ satisfying (10), (11) and (12), such that

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^{n} w_i x_{(i)}\}, \ \forall \tilde{x} \in \mathbb{R}^n. \tag{14}$$

Proof. See the on-line supplement. □

By Theorems 3 and 4, we see the main differences between natural risk statistics and coherent risk measures:

(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 a weighted sample average. There is no simple linear function that can transform a L-statistic to a 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 reasonable set of axioms.

(3) A law-invariant coherent risk statistic is a supremum of L-statistic with increasing weights. Hence, if one assigns larger weights to larger observations, a natural risk statistic become a law invariant coherent risk statistic. However, assigning larger weights to larger observations is not robust.

7.2 Comparison with Insurance Risk Measures

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

Definition 3. A risk statistic $\hat{\rho} : \mathbb{R}^n \to \mathbb{R}$ is called a insurance risk statistic, if it satisfies the following axioms F1-F4.

Axiom F1. Permutation invariance: the same as Axiom C4.

Axiom F2. Monotonicity: $\hat{\rho}(\tilde{x}) \leq \hat{\rho}(\tilde{y})$, if $\tilde{x} \leq \tilde{y}$.

Axiom F3. 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 F4. Scale normalization: $\hat{\rho}(\mathbf{1}) = 1$.

We have the following representation theorem for the insurance risk statistic.

Theorem 5. Let $x_{(1)}, ..., x_{(n)}$ be the order statistics of the observation $\tilde{x}$ with $x_{(n)}$ being the largest, then $\hat{\rho}$ is an insurance risk statistic if and only if there exists a single weight $\tilde{w} = (w_1, \ldots, w_n)$ with $w_i \geq 0$ for $i = 1, \ldots, n$ and $\sum_{i=1}^n w_i = 1$, such that

$$\hat{\rho}(\tilde{x}) = \sum_{i=1}^{n} w_i x_{(i)}, \ \forall \tilde{x} \in \mathbb{R}^n. \tag{15}$$

Proof. See the on-line supplement. □

Comparing (6) and (15), we see that a natural risk statistic is the supremum of L-statistics, while an insurance risk statistic is just one L-statistic. Therefore, insurance risk statistics cannot incorporate different scenarios. On the other hand, each weight $\tilde{w} = (w_1, \ldots, w_n)$ 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, nature risk statistics incorporate the idea of evaluating the risk under different scenarios, similar to coherent risk measures.

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

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

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 with weights being $(0.5, 0.5, 0)$ and $(0.72, 0.08, 0.2)$.

8 Tail Conditional Median: a Robust Natural Risk Statistic

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

8.1 The Differences between Tail Conditional Expectation and Tail Conditional Median

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 α is defined as

$$\text{TCM}_\alpha(X) = \text{median}[X | X \geq \text{VaR}_\alpha(X)]. \tag{16}$$

In other words $\mathrm{TCM}_\alpha(X)$ is the conditional median of X given that $X \geq \mathrm{VaR}_\alpha(X)$.

Remark: If X is continuous then

$$\mathrm{TCM}_\alpha(X) = \mathrm{VaR}_{\frac{1+\alpha}{2}}(X).$$

This show that VaR at a higher level can incorporate tail information, contrary to some claims in the existing literature. For example, if one wants to measure the loss beyond 95% level, one can use VaR at 97.5%, which is the tail conditional median at 95% level. For discrete random variable or data, one simply uses the definition (16) and there may be a difference between $\mathrm{TCM}_\alpha(X)$ and $\mathrm{VaR}_{\frac{1+\alpha}{2}}(X)$, depending on ways of defining quantiles for discrete data.

There have been some examples in the existing literature that are used to show VaR does not satisfy subadditivity at certain level α. However, if one considers TCM at the same level α, or equivalently considers VaR at a higher level, the problem of non-subadditivity of VaR is easily solved. We list some major examples here:

Example 1. The example on page 216 of [3] did not calculate VaR correctly. Actually in that example, the 1% VaR[13] of two options A and two options B are $2u$ and $2l$ respectively, instead of $-2u$ and $-2l$. And the 1% VaR of $A+B$ is $u+l$, instead of $100-l-u$. Therefore, VaR satisfies subadditivity in that example.

Example 2. The example on page 217 of [3] showed that the 10% VaR does not satisfy subadditivity for X_1 and X_2. However, the 10% tail conditinal median (or equivalently 5% VaR) satisfies subadditivity! Actually, the 5% VaR of X_1 and X_2 are both equal to 1. By simple calculation, $P(X_1 + X_2 \leq -2) = 0.005 < 0.05$, which implies that the 5% VaR of $X_1 + X_2$ is strictly less than 2.

Example 3. The example in section 2.1 of [20] showed that the 99% VaR of L_1 and L_2 are equal, although apparently L_2 is much more risky than L_1. However, the tail conditional median at 99% level (or 99.5% VaR), of L_1 is equal to 10^{10}, which is much larger than 1, the tail conditional median at 99% level (99.5% VaR) of L_2. In other words, if one looks at the tail conditional median at 99% level, one can correctly compare the risk of the two portfolios.

There are several differences between TCE and TCM. First, there are theoretical differences. For example, TCM does not in general satisfy subadditivity, although TCE generally

[13] In [3], VaR is defined as $\mathrm{VaR}(X) = -\inf\{x \mid P(X \leq x) > \alpha\}$, where $X = -L$ representing the net worth of a position. In other words, VaR at level α in [3] corresponds to VaR at level $1-\alpha$ in this paper.

does; and TCM is a natural risk statistic while TCE is a coherent risk statistic.

Second, as we shall see, theoretically TCM is more robust than TCE because TCM has a bounded influence function but TCE does not.

Third, there may be significant numerical differences between TCE and TCM. In Table 1, we calculated the risk measure TCE_α and TCM_α with α ranging from 95% to 99% for a data set of auto insurance claims. Both the differences and relative differences between TCE_α and TCM_α are very significant. Table 2 uses S&P 500 daily data from January 03, 1980 to December 21, 2005. More precisely, we report TCE and TCM for the daily losses (negative returns) in Table 2 with α ranging from 95.0% to 99.9%. The relative differences of TCE and TCM are also very significant.

α	TCE_α	TCM_α	$\mathrm{TCE}_\alpha - \mathrm{TCM}_\alpha$	$\frac{\mathrm{TCE}_\alpha - \mathrm{TCM}_\alpha}{\mathrm{TCE}_\alpha}$
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 1: The difference between tail conditional expectation (TCE) and tail conditional median (TCM) for a data set of auto insurance claim. The table shows a significant difference between the TCE and TCM.

8.2 Robustness Comparison between the Tail Conditional Expectation and Tail Conditional Median

Next we show numerically that the tail conditional median is more robust than the tail conditional expectation. The left panel of Figure 1 shows the value of TCE_α with respective to $\log(1-\alpha)$ for Laplace distribution and T-distribution, where α is in the range $[0.95, 0.999]$. As demonstrated in Figure 1, if the model assumes the loss distribution to be Laplace while the underlying true loss distribution is a t-distribution, the calculated TCE value can be far from the true value. The right panel of Figure 1 shows the value of TCM_α with respective to $\log(1-\alpha)$ for Laplace distribution and T-distribution. As seen from the figure, TCM_α is more robust than TCE_α in the sense that it is less sensitive to the tail behavior of the

α	TCE_α	TCM_α	$\mathrm{TCE}_\alpha - \mathrm{TCM}_\alpha$	$\frac{\mathrm{TCE}_\alpha - \mathrm{TCM}_\alpha}{\mathrm{TCE}_\alpha}$
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 2: The tail conditional expectation (TCE) and tail conditional median (TCM) for S&P 500 index daily losses (negative returns) from Jan 03, 1980 to Dec 21, 2005. The table shows a significant difference between the TCE and TCM.

underlying distribution. For example, as shown in the figure, with $\alpha = 99.6\%$, the variation of TCE with respect to the change of underlying models is 1.44, whereas the variation of TCM is only 0.75.

8.3 Influence Functions of Tail Conditional Expectation and Tail Conditional Median

Influence functions introduced by Hampel [30] are useful in assessing the robustness of an estimator. Consider an estimator $T(F)$ based on an unknown distribution F. For $x \in \mathbb{R}$, let δ_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.

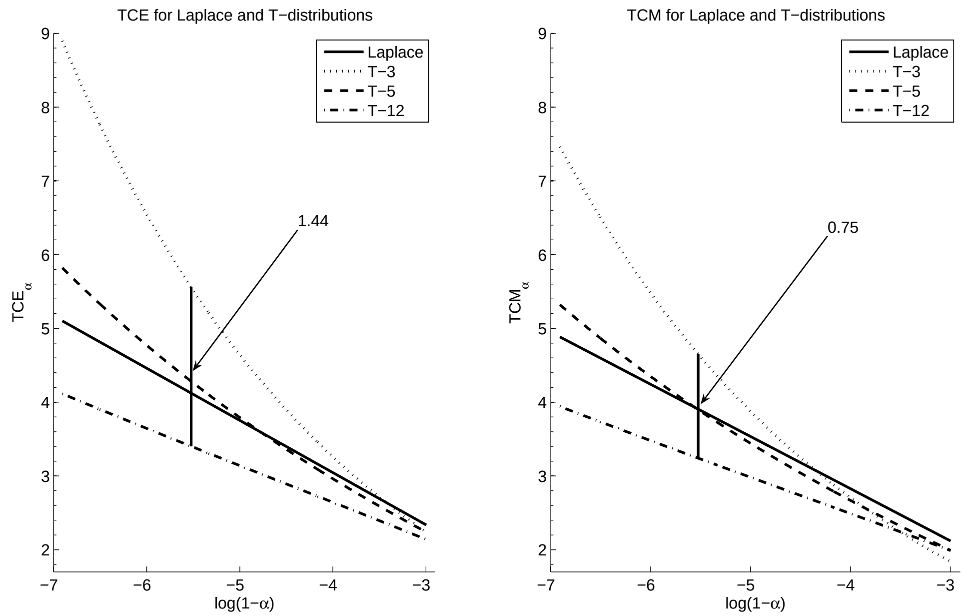

Figure 1: Comparison of the robustness of the tail contional expectation (TCE) and tail conditional median (TCM). The distributions used are Laplace and T-distributions with degree of freedom 3, 5, 12, normalized to have mean 0 and variance 1. The x-axis is $\log(1-\alpha)$ where $\alpha \in [0.95, 0.999]$. The tail conditional median (TCM) is less sensitive to changes in distribution, as the right panel has a narrower range in y-axis.

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

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

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

$$IF(x, \mathrm{TCE}_\alpha, X) = \begin{cases} \mathrm{VaR}_\alpha(X) - E[X|X \geq \mathrm{VaR}_\alpha(X)], & \text{if } x \leq \mathrm{VaR}_\alpha(X), \\ \frac{x}{1-\alpha} - E[X|X \geq \mathrm{VaR}_\alpha(X)] - \frac{\alpha}{1-\alpha}\mathrm{VaR}_\alpha(X), & \text{if } x > \mathrm{VaR}_\alpha(X). \end{cases} \quad (17)$$

Proof. See the on-line supplement. □

We see immediately from Proposition 1 that

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

Hence, TCE has an unbounded influence function but TCM has a bounded influence function, which implies that TCM is more robust.

8.4 Discussion on Computational Issues

There are at least two computational issues: whether it is easy to compute a risk measure from the regulator's viewpoint, and whether it is easy to incorporate a risk measure into portfolio optimization from an individual bank's viewpoint.

For the first issue, since the tail conditional median is robust, it is easier to compute the tail conditional median than tail conditional expectation, as the tail conditional median is less sensitive to modelling assumptions.

For the second issue, it is easier to do portfolio optimization with respect to the tail conditional expectation than to the tail conditional median, as the mean leads to convexity in optimization. However, we should point out that doing optimization with respect to median is a classical problem in robust statistics, and recently there are good algorithms designed for portfolio optimization under both CVaR and VaR constraints (see [45]). Furthermore, from the regulator's viewpoint, it is a first priority to find a good robust risk measure for the purpose of legal implementation. How to achieve better profits via portfolio optimization, under the risk measure constraints to satisfy governmental regulations, should be a matter left for individual banks.

9 Conclusion

We propose new, data-based, risk measures, called natural risk statistics, 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 coherent risk measures, and relaxing the comonotonic additivity in insurance risk measures. The relaxation is consistent with the prospect theory in psychology. Comparing to previous risk measures, the natural risk statistics include the tail conditional median which is more robust than the tail conditional expectation suggested by coherent risk measures; and, unlike the insurance risk measure, the natural risk statistics can also incorporate scenario analysis. The natural risk statistics include VaR (with senario analysis) as a special case and therefore shows that VaR, though simple, is not irrational.

We emphasize that the objectives of risk measures are very relevant for our discussion. In particular, some risk measures may be suitable for internal management but not for external regulatory agencies, and vice versa. For example, coherent and convex risk measures may be good for internal risk measures, as there are connections between these risk measures and subjective prices in incomplete markets for market makers (see, e.g., the connections between coherent and convex risk measures and good deal bounds in Jaschke and Küchler [36] and Staum [56]). However, as we point out, for external risk measures one may prefer a different set of properties, including consistency in implementation which means robustness.

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

Acknowledgements: We thank many people who offered insight into this work, including J. Birge, L. Eeckhoudt, M. Frittelli, and J. Staum. This research is supported in part by National Science Foundation.

References

[1] Acerbi, C., 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.

[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] Artzner, P., F. Delbaen, J.-M. Eber, D. Heath and H. Ku (2003). Coherent multiperiod risk adjusted values and Bellman's principle. Preprint.

[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] Bion-Nadal, J. (2004). Conditional risk measures and robust representation of convex risk measures. Ecole Polytechnique, CMAP, preprint no. 557.

[9] Castagnoli, E., G. Favero and Maccheroni (2006) A problem in sublinear pricing along time. Working paper, Univerita Commerciale Luigi Bocconi, Milano, Italy.

[10] Cheridito, P., F. Delbaen and M. Kupper (2005). Coherent and convex monetary risk measures for unbounded càdlàg processes. *Finance and Stochastics*, **9(3)**, 369-387.

[11] Cheridito, P., F. Delbaen and M. Kupper (2006). Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, **11**, 57-106.

[12] Cheridito, P. and M. Kupper (2006) Composition of time-consistent dynamic monetary risk measures in discrete time. Preprint, Princeton University.

[13] Daníelsson, J., B. N. Jorgensen, G. Samorodnitsky, M. Sarma and 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.

[14] Dana, R.-A. (2005). A representation result for concave Schur concave functions. *Mathematical Finance*, **15**, 615-634.

[15] Delbaen, F. (2002). Coherent risk measures on general probability spaces. *Advances in Finance and Stochastics–Essays in Honour of Dieter Sondermann*, K. Sandmann and P. J. Schönbucher eds., Springer, New York.

[16] Denneberg, D. (1994). *Non-Additive Measure and Integral*, Kluwer Academic Publishers, Boston.

[17] Detlefsen, K. and G. Scandolo (2005). Conditional and dynamic convex risk measures. *Finance and Stochastics*, **9**, 539-561.

[18] Dhaene, J., M. J. Goovaerts and R. Kaas (2003). Economic capital allocation derived from risk measures. *North American Actuarial Journal*, **7**, 44-56.

[19] Dhaene, J., S. Vanduffel, M.J. Goovaerts, R. Kaas, Q. Tang and D. Vyncke (2006). Risk measures and comonotonicity: a review. *Stochastic Models*, **22**, 573-606.

[20] Dunkel, J. and Weber, S. Efficient Monte Carlo Methods for Convex Risk Measures. Preprint. 2007.

[21] Eeckhoudt, L. and H. Schlesinger (2006). Putting risk in its proper place. *American Economic Review*, **96** 1, 280-289.

[22] 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.

[23] Föllmer, H. and A. Schied (2002). Convex measures of risk and trading constraints, *Finance and Stochastics*, **6**, 429-447.

[24] Föllmer, H. and A. Schied (2004). *Stochastic Finance, An Introduction in Discrete Time*, 2nd revised and extended edition. Welter de Gruyter, Berlin, New York.

[25] Föllmer, H. and I. Penner (2006). Convex risk measures and the dynamics of their penalty functions. Preprint, Humboldt-University Berlin.

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

[27] Frittelli, M. and E. R. Gianin(2005). Law invariant convex risk measures, *Advances in Mathematical Economics*, **7**, 33-46.

[28] Garcia, R., E. Renault and G. Tsafack (2007). Proper conditioning for coherent VaR in portfolio management. *Management Science*, **53**, 483-494.

[29] Hart, H. L. A. (1994). *The Concept of Law*, 2nd, Clarendon Press, Oxford.

[30] Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of American Statistical Association*, **62**, 1179-1186.

[31] Hennessy, D. A. and Lapan, H. (2006). On the nature of certainty equivalent functionals. *Journal of Mathematical Economics*, **43**, 1-10.

[32] Heyde, C. C. and S. G. Kou (2004). On the controversy over tailweight of distributions. *Operations Research Letters*, **32**, 399-408.

[33] Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

[34] Ibragimov, R. and Walden, J. (2006). The limits of diversification when losses may be large. Harvard Institute of Economic Research Discussion Paper No. 2104. Available at SSRN: http://ssrn.com/abstract=880081.

[35] Institute of Transportation Engineers, Speeding zoning information, available at http://www.ite.org/standards/speed_zoning.pdf.

[36] Jaschke, S., and U. Küchler (2001). Coherent risk measures and good deal bounds. *Finance and Stochastics*. **5**, 181-200.

[37] Jobert, A. and L.C.G. Rogers (2005). Valuations and dynamic convex risk measures. Preprint, Cambridge University.

[38] Kahneman, D. and A. Tversky (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, **47**, 1979, 263-292.

[39] Kimball, M. (1993). Standard risk aversion. *Econometrika*, **61**, 589-611.

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

[41] Peng, S. (2004). Non-linear expectations, non-linear evaluations and risk measures. *Lecture Notes in Mathematics*, **1856**, Springer.

[42] 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 Academic Publishers (Norwell 2000), 278-287.

[43] Quiggin, J. (1982). A theory of anticipated utility. *Journal of economic behavior and organization*, **3**, 323-343.

[44] Riedel, F. (2004). Dynamic coherent risk measures. *Stochastic Processes and their Applications*, **112**, 185-200.

[45] Rockafellar, R.T. and S. Uryasev (2000). Optimization of Conditional Value-At-Risk. *Journal of Risk*, **2**, 21-41.

[46] Rockafellar, R. T. and S. Uryasev (2002). Conditional Value-at-Risk for general loss distributions. *Journal of Banking and Finance*, **26**, 1443-1471.

[47] Roorda, B., J.M. Schumacher, J. Engwerda (2005). Coherent acceptability measures in multi-period models. *Math. Finance*, **15**, 589-612.

[48] Rüschendorf, L. (2005) Law invariant convex risk measures for portfolio vectors. Preprint, University of Freiburg.

[49] Samuelson, P. A. (1967). General proof that diversification pays. *The Journal of Financial and Quantitative Analysis*, **2**, 2-13.

[50] Schied, A. (2006). Risk measures and robust optimization problems. *Stochastic Models*, **22**, 753-831.

[51] Schmeidler, D. (1986). Integral representation without additivity. *Proceedings of the American Mathematical Society*, **97**, 255-261.

[52] Schmeidler, D. (1989). Subjective probability and expected utility without additivity, *Econometrica*, **57**, 571-587.

[53] Song, Y. and J.-A. Yan (2006). The representations of two types of functionals on $L^\infty(\Omega, \mathcal{F})$ and $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. *Science in China Series A: Mathematics*, **49**, 1376-1382.

[54] Song, Y. and J.-A. Yan (2006). Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. Preprint, Chinese Academy of Sciences.

[55] Staudte, R. G., and S. J. Sheather (1990). *Robust Estimation and Testing*, Wiley, New York.

[56] Staum, J. (2004). Fundamental theorems of asset pricing for good deal bounds. *Mathematical Finance.* **14**, 141–161.

[57] Transportation Research Board of National Academies (2003). *Design Speed, Operating Speed, and Posted Speed Practices (NCHRP report 504)*, available at http://trb.org/publications/nchrp/nchrp_rpt_504.pdf.

[58] Tversky, A. and D. Kahneman (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**, 297-323.

[59] Wakker, P. and A. Tversky (1993). An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty*, **7**, 147-176.

[60] Wang, S. S., V. R. Young and H. H. Panjer (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, **21**, 173-183.

[61] Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance*, **16(2)**, 419-441.

[62] Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**, 95-115.

On-Line Supplement

What Is a Good Risk Measure: Bridging the Gaps between Robustness, Subadditivity, and Insurance Risk Measures

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

Columbia University

1 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 convergence.

Lemma 1. Let $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \leq y_2 \leq \cdots \leq y_n\}$, and denote $\mathcal{B}^o$ to be the interior of $\mathcal{B}$. For any fixed $\tilde{z} = (z_1, \ldots, z_n) \in \mathcal{B}^o$ and any $\hat{\rho}$ satisfying axioms C1-C4 and $\hat{\rho}(\tilde{z}) = 1$ there exists a weight $\tilde{w} = (w_1, \ldots, w_n)$ such that the linear functional $\lambda(\tilde{x}) := \sum_{i=1}^n w_i x_i$ satisfies

$$\lambda(\tilde{z}) = 1, \tag{18}$$

$$\lambda(\tilde{x}) < 1 \text{ for all } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x}) < 1. \tag{19}$$

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 C1 and C3 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 $\tilde{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n$ 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

$$\lambda(\tilde{x}) \leq \lambda(\tilde{z}), \forall \tilde{x} \in U. \tag{20}$$

We shall show that the strict inequality holds in (20). 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

$$\lambda(\tilde{x}^\alpha) = \alpha\lambda(\tilde{z}) + (1 - \alpha)\lambda(\tilde{x}^0) = \lambda(\tilde{z}) \tag{21}$$

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

$$\hat{\rho}(\tilde{x}^\alpha) \leq \alpha\hat{\rho}(\tilde{z}) + (1-\alpha)\hat{\rho}(\tilde{x}^0) < \alpha + (1-\alpha) = 1, \quad \forall \alpha \in (0,1). \tag{22}$$

Since $\tilde{z} \in \mathcal{B}^o$, 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 $\varepsilon > 0$,

$$\tilde{x}^{\alpha_0} + \varepsilon\tilde{w} \in \mathcal{B}. \tag{23}$$

With $w_{\max} = \max(w_1, w_2, ..., w_n)$, we have $\tilde{x}^{\alpha_0} + \varepsilon\tilde{w} \leq \tilde{x}^{\alpha_0} + \varepsilon w_{\max}\mathbf{1}$. Thus, the monotonicity in Axiom C2 and translation invariance in Axiom C1 yield

$$\hat{\rho}(\tilde{x}^{\alpha_0} + \varepsilon\tilde{w}) \leq \hat{\rho}(\tilde{x}^{\alpha_0} + \varepsilon w_{\max}\mathbf{1}) = \hat{\rho}(\tilde{x}^{\alpha_0}) + \varepsilon w_{\max}. \tag{24}$$

Since $\hat{\rho}(\tilde{x}^{\alpha_0}) < 1$ via (22), we have by (24) and (23) 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, (20) implies $\lambda(\tilde{x}^{\alpha_0} + \varepsilon\tilde{w}) \leq \lambda(\tilde{z})$. However, we have, by (21), 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

$$\lambda(\tilde{x}) < \lambda(\tilde{z}), \forall \tilde{x} \in U. \tag{25}$$

Since $\hat{\rho}(0) = 0$, we have $0 \in U$. Letting $\tilde{x} = 0$ in (25) yields $\lambda(\tilde{z}) > 0$, so we can re-scale $\tilde{w}$ such that $\lambda(\tilde{z}) = 1 = \hat{\rho}(\tilde{z})$. Thus, (25) 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 (19) holds. □

Lemma 2. Let $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \leq y_2 \leq \cdots \leq y_n\}$, and denote $\mathcal{B}^o$ to be the interior of $\mathcal{B}$. For any fixed $\tilde{z} = (z_1, \ldots, z_n) \in \mathcal{B}^o$ and any $\hat{\rho}$ satisfying axioms C1-C4, there exists a weight $\tilde{w} = (w_1, \ldots, w_n)$ such that

$$\sum_{i=1}^n w_i = 1, \tag{26}$$

$$w_i \geq 0, i = 1, \ldots, n, \tag{27}$$

$$\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^n w_i x_i, \text{ for } \forall \tilde{x} \in \mathcal{B}, \text{ and } \hat{\rho}(\tilde{z}) = \sum_{i=1}^n w_i z_i. \tag{28}$$

Proof. We will show this by considering three cases.

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

From Lemma 1, there exists a weight $\tilde{w} = (w_1, \ldots, w_n)$ such that the linear functional $\lambda(\tilde{x}) := \sum_{i=1}^n w_i x_i$ satisfies (18) and (19).

First we prove that $\tilde{w}$ satisfies (26). 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 C1 implies $\hat{\rho}(c\mathbf{1}) = c < 1$. Thus, (19) implies $\lambda(c\mathbf{1}) < 1$, and, by continuity of λ, we obtain that $\lambda(\mathbf{1}) \leq 1$. Secondly, for any $c > 1$, Axiom C1 implies $\hat{\rho}(2\tilde{z} - c\mathbf{1}) = 2\hat{\rho}(\tilde{z}) - c = 2 - c < 1$. Then it follows from (19) and (18) 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 (26).

Next, we will prove that $\tilde{w}$ satisfies (27). Let $\tilde{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0)$ 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 C2. Then (19) and (18) 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 ε go to 0.

Finally, we will prove that $\tilde{w}$ satisfies (28). It follows from Axiom C1 and (19) that

$$\forall c > 0,\ \lambda(\tilde{x}) < c,\ \text{for all } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x}) < c. \tag{29}$$

For any $c \leq 0$, we choose $b > 0$ such that $b + c > 0$. Then by (29), we have

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

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

$$\forall c \leq 0,\ \lambda(\tilde{x}) < c,\ \text{for all } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x}) < c. \tag{30}$$

It follows from (29) and (30) that

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

which in combination with $\hat{\rho}(\tilde{z}) = \lambda(\tilde{z}) = 1$ completes the proof of (28).

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 $\tilde{w} = (w_1, \ldots, w_n)$ satisfying (26), (27) 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 (28).

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 $\tilde{w} = (w_1, \ldots, w_n)$ satisfying (26), (27) 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 (28). □

Proof of Theorem 1. (1) The proof of part (I). Suppose $\hat{\rho}$ is defined by (5), then obviously $\hat{\rho}$ satisfies axioms C1 and C4.

To check Axiom C2, write

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

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

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

which implies that $\hat{\rho}$ satisfies Axiom C2 because

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

To check Axiom C3, note that if $\tilde{x}$ and $\tilde{y}$ are comonotonic, then there exists a permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$ such that $x_{i_1} \leq x_{i_2} \leq \ldots \leq x_{i_n}$ and $y_{i_1} \leq y_{i_2} \leq \ldots \leq y_{i_n}$.

Hence, we have $(\tilde{x}+\tilde{y})_{(i)} = x_{(i)} + y_{(i)}$, $i = 1, ..., n$. Therefore,

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

which implies that $\hat{\rho}$ satisfies Axiom C3.

(2) The proof of part (II). By Axiom C4, we only need to show that there exists a set of weights $\mathcal{W} = \{\tilde{w} = (w_1, \ldots, w_n)\} \subset \mathbb{R}^n$ 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

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

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

By Lemma 2, for any point $\tilde{y} \in \mathcal{B}^o$, there exists a weight $\tilde{w}^{(\tilde{y})}$ satisfying (26), (27) and (28). Therefore, we can take the collection of such weights as

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

Then from (28), for any fixed $\tilde{x} \in \mathcal{B}^o$ we have

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

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

Therefore,

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{y}\in\mathcal{B}^o} \{\sum_{i=1}^{n} w_i^{(\tilde{y})} x_i\} = \sup_{\tilde{w}\in\mathcal{W}} \{\sum_{i=1}^{n} w_i x_i\}, \ \forall \tilde{x} \in \mathcal{B}^o, \tag{31}$$

where each $\tilde{w} \in \mathcal{W}$ satisfies (26) and (27).

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

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w}\in\mathcal{W}} \{\sum_{i=1}^{n} w_i x_i\}, \ \forall \tilde{x} \in \partial\mathcal{B}. \tag{32}$$

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

$$\hat{\rho}(\tilde{x}^0) = \lim_{k\to\infty} \hat{\rho}(\tilde{x}^k) = \lim_{k\to\infty} \sup_{\tilde{w}\in\mathcal{W}} \{\sum_{i=1}^{n} w_i x_i^k\}, \tag{33}$$

where the last equality follows from (31). If we can interchange sup and limit in (33), i.e. if

$$\lim_{k\to\infty} \sup_{\tilde{w}\in\mathcal{W}} \{\sum_{i=1}^{n} w_i x_i^k\} = \sup_{\tilde{w}\in\mathcal{W}} \{\lim_{k\to\infty} \sum_{i=1}^{n} w_i x_i^k\} = \sup_{\tilde{w}\in\mathcal{W}} \{\sum_{i=1}^{n} w_i x_i^0\}, \tag{34}$$

then (32) holds and the proof is complete.

To show (34), note that we have by Cauchy-Schwartz inequality

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

because $w_i \ge 0$ and $\sum_{i=1}^{n} w_i = 1, \forall \tilde{w} \in \mathcal{W}$. Therefore, $\sum_{i=1}^{n} w_i x_i^k \to \sum_{i=1}^{n} w_i x_i^0$ uniformly for all $\tilde{w} \in \mathcal{W}$ and (34) follows. $\square$

2 Proof of Theorem 2

(I) (1) For $\forall \tilde{x} \le 0$, Axiom C2 implies $\hat{\rho}(\tilde{x}) \le \hat{\rho}(0) = 0$, hence $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$ by definition. Thus, D1 holds. (2) For any $\tilde{x} \in \mathbb{R}^n_{++}$, there exists $\alpha > 0$ such that $0 \le \tilde{x} - \alpha\mathbf{1}$. Axioms C2 and C1 imply that $\hat{\rho}(0) \le \hat{\rho}(\tilde{x} - \alpha\mathbf{1}) = \hat{\rho}(\tilde{x}) - \alpha$. So $\hat{\rho}(\tilde{x}) \ge \alpha > 0$ and henceforth $\tilde{x} \notin \mathcal{A}_{\hat{\rho}}$, i.e., D2 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 $\hat{\rho}(\tilde{x}) \le 0$, $\hat{\rho}(\tilde{y}) \le 0$, and $\lambda\tilde{x}$ and $(1-\lambda)\tilde{y}$ are comonotonic for any $\lambda \in [0,1]$. Thus C3 implies

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

Hence $\lambda\tilde{x}+(1-\lambda)\tilde{y} \in \mathcal{A}_{\hat{\rho}}$, i.e., D3 holds. (4) For any $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$ and $a > 0$, we have $\hat{\rho}(\tilde{x}) \le 0$ and C1 implies $\hat{\rho}(a\tilde{x}) = a\hat{\rho}(\tilde{x}) \le 0$. Thus, $a\tilde{x} \in \mathcal{A}_{\hat{\rho}}$, i.e., D4 holds. (5) For any $\tilde{x} \le \tilde{y}$ and $\tilde{y} \in \mathcal{A}_{\hat{\rho}}$, we have $\hat{\rho}(\tilde{y}) \le 0$. By C2, $\hat{\rho}(\tilde{x}) \le \hat{\rho}(\tilde{y}) \le 0$. Hence $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$, i.e., D5 holds. (6) If $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$, then $\hat{\rho}(\tilde{x}) \le 0$. For any permutation $(i_1, \ldots, i_n)$, C4 implies $\hat{\rho}((x_{i_1}, \ldots, x_{i_n})) = \hat{\rho}(\tilde{x}) \le 0$. So $(x_{i_1}, \ldots, x_{i_n}) \in \mathcal{A}_{\hat{\rho}}$, i.e., D6 holds. (7) Suppose $\tilde{x}^k \in \mathcal{A}_{\hat{\rho}}, k = 1, 2, \ldots$, and $\tilde{x}^k \to \tilde{x}$ as $k \to \infty$. Then $\hat{\rho}(\tilde{x}^k) \le 0, \forall k$. 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 $\tilde{x}^k \to \tilde{x}$, it follows that there exists $K \in \mathbb{N}$ such that $\tilde{x}^K > \tilde{x} - \delta\mathbf{1}$. By C2, $\hat{\rho}(\tilde{x}^K) \ge \hat{\rho}(\tilde{x} - \delta\mathbf{1}) > 0$, which contradicts to $\hat{\rho}(\tilde{x}^K) \le 0$. 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 $\forall \tilde{x} \in \mathbb{R}^n, \forall a \ge 0$, 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 D1 and D2. 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 D4. Therefore, C1 holds. (2) Suppose $\tilde{x} \leq \tilde{y}$. For any $m \in \mathbb{R}$, if $\tilde{y} - m\mathbf{1} \in \mathcal{A}$, then D5 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., C2 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 D3 that $\frac{1}{2}(\tilde{x} - m\mathbf{1}) + \frac{1}{2}(\tilde{y} - n\mathbf{1}) \in \mathcal{A}$. By D4, 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 C3 holds. (4) Fix any $\tilde{x} \in \mathbb{R}^n$ and any permutation $(i_1, \ldots, i_n)$. Then for any $m \in \mathbb{R}$, D6 implies that $\tilde{x} - m\mathbf{1} \in \mathcal{A}$ if and only if $(x_{i_1}, \ldots, x_{i_n}) - m\mathbf{1} \in \mathcal{A}$. Hence

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

Taking infimum on both sides, we obtain $\hat{\rho}_{\mathcal{A}}(\tilde{x}) = \hat{\rho}_{\mathcal{A}}((x_{i_1}, \ldots, x_{i_n}))$, i.e., D.4 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 C1. On the other hand, for any $\delta > \hat{\rho}(\tilde{x})$, we have

$$\begin{aligned}\delta > \hat{\rho}(\tilde{x}) &\Rightarrow \hat{\rho}(\tilde{x} - \delta\mathbf{1}) < 0 \text{ (since } \hat{\rho} \text{ satisfies C1)} \\ &\Rightarrow \tilde{x} - \delta\mathbf{1} \in \mathcal{A}_{\hat{\rho}} \text{ (by definition)} \\ &\Rightarrow \hat{\rho}_{\mathcal{A}_{\hat{\rho}}}(\tilde{x} - \delta\mathbf{1}) \leq 0 \text{ (by definition)} \\ &\Rightarrow \hat{\rho}_{\mathcal{A}_{\hat{\rho}}}(\tilde{x}) \leq \delta \text{ (since } \hat{\rho}_{\mathcal{A}_{\hat{\rho}}} \text{ satisfies C1)}\end{aligned}$$

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 D5 $\tilde{x} \in \mathcal{D}$. 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 $k \to \infty$ 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 $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}} = \bar{\mathcal{D}}$. □

3 Proof of Theorem 3

The proof for the "if" part is trivial. The proof for the "only if" part. Suppose $\hat{\rho}$ is a coherent risk statistic. Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ be an arbitrary set with n elements. Let $\mathcal{H}$ be the set of all the subsets of Θ. Let $\mathcal{Z}$ be the set of all real-valued random variables defined on $(\Theta, \mathcal{H})$. Define a functional E^* on $\mathcal{Z}$:

$$\begin{aligned} E^* &: \mathcal{Z} \to \mathbb{R} \\ & Z \mapsto E^*(Z) \triangleq \hat{\rho}(Z(\theta_1), Z(\theta_2), \ldots, Z(\theta_n)), \end{aligned}$$

then $E^*(\cdot)$ is monotone, positively affinely homogeneous and subadditive in the sense defined in equations (2.7), (2.8) and (2.9) at Chapter 10 of Huber [33]. Then the result follows by applying Proposition 2.1 at page 254 of Huber [33] to E^*. □

4 Proof of Theorem 4

The proof for theorem 4 follows the same line as the proof for theorem 1. We first prove two lemmas which are similar to Lemma 1 and 2, with simpler argument and stronger conclusion. More precisely, we have the following two lemmas.

Lemma 3. Let $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \leq y_2 \leq \cdots \leq y_n\}$. For any fixed $\tilde{z} \in \mathcal{B}$ and any $\hat{\rho}$ satisfying axioms C1, C2, E3, C4 and $\hat{\rho}(\tilde{z}) = 1$, there exists a weight $\tilde{w} = (w_1, \ldots, w_n)$ satisfying (12) such that the linear functional $\lambda(\tilde{x}) := \sum_{i=1}^n w_i x_i$ satisfies

$$\lambda(\tilde{z}) = 1, \tag{35}$$

$$\lambda(\tilde{x}) < 1 \text{ for all } \tilde{x} \text{ such that } \hat{\rho}(\tilde{x}) < 1. \tag{36}$$

Proof. Let $U = \{\tilde{x} \mid \hat{\rho}(\tilde{x}) < 1\}$. Axioms C1 and E3 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 $\tilde{w}^0 = (w_1^0, \ldots, w_n^0) \in \mathbb{R}^n$ such that $\lambda^0(\tilde{x}) := \sum_{i=1}^n w_i^0 x_i$ satisfies $\lambda^0(\tilde{x}) \leq \lambda^0(\tilde{z})$ for all $\tilde{x} \in \bar{U}$. In particular, we have

$$\lambda^0(\tilde{x}) \leq \lambda^0(\tilde{z}), \forall \tilde{x} \in U. \tag{37}$$

Let $(i_1, \ldots, i_n)$ be the permutation of $(1, 2, \ldots, n)$ such that $w_{i_1}^0 \leq w_{i_2}^0 \leq \cdots \leq w_{i_n}^0$. And let $(j_1, \ldots, j_n)$ be the permutation of $(1, 2, \ldots, n)$ such that $i_{j_k} = k, k = 1, 2, \ldots, n$. Define a new weight $\tilde{w}$ and a new linear functional as follows:

$$\tilde{w} = (w_1, \ldots, w_n) \triangleq (w_{i_1}^0, \ldots, w_{i_n}^0), \tag{38}$$

$$\lambda(\tilde{x}) := \sum_{i=1}^n w_i x_i, \tag{39}$$

then $\tilde{w}$ satisfies condition (12).

For any fixed $\tilde{x} \in U$, by Axiom C4, $\hat{\rho}((x_{j_1}, \ldots, x_{j_n})) = \hat{\rho}(\tilde{x}) < 1$, so $(x_{j_1}, \ldots, x_{j_n}) \in U$. Then, we have

$$\begin{aligned}
\lambda(\tilde{x}) &= \sum_{k=1}^n w_k x_k = \sum_{k=1}^n w_{i_k}^0 x_k = \sum_{k=1}^n w_{i_{j_k}}^0 x_{j_k} \\
&= \sum_{k=1}^n w_k^0 x_{j_k} = \lambda^0((x_{j_1}, \ldots, x_{j_n})) \leq \lambda^0(\tilde{z})
\end{aligned} \tag{40}$$

where the last inequality follows from (37). Noting that $z_1 \leq \ldots \leq z_n$, we obtain

$$\lambda^0(\tilde{z}) = \sum_{k=1}^n w_k^0 z_k \leq \sum_{k=1}^n w_{i_k}^0 z_k = \lambda(\tilde{z}). \tag{41}$$

By (40) and (41), we have

$$\lambda(\tilde{x}) \leq \lambda(\tilde{z}), \forall \tilde{x} \in U. \tag{42}$$

We shall show that the strict inequality holds in (42). Suppose, by contradiction, that there exists $\tilde{x}^0 \in U$ such that $\lambda(\tilde{x}^0) = \lambda(\tilde{z})$. With $w_{\max} = \max(w_1, w_2, ..., w_n)$, we have $\tilde{x}^0 + \varepsilon\tilde{w} \leq \tilde{x}^0 + \varepsilon w_{\max}\mathbf{1}$ for any $\varepsilon > 0$. Thus, axioms C1 and C2 yield

$$\hat{\rho}(\tilde{x}^0 + \varepsilon\tilde{w}) \leq \hat{\rho}(\tilde{x}^0 + \varepsilon w_{\max}\mathbf{1}) = \hat{\rho}(\tilde{x}^0) + \varepsilon w_{\max}, \ \forall \varepsilon > 0. \tag{43}$$

Since $\hat{\rho}(\tilde{x}^0) < 1$, we have by (43) that for small enough $\varepsilon > 0$, $\hat{\rho}(\tilde{x}^0 + \varepsilon\tilde{w}) < 1$. Hence, $\tilde{x}^0 + \varepsilon\tilde{w} \in U$ and (42) implies $\lambda(\tilde{x}^0 + \varepsilon\tilde{w}) \leq \lambda(\tilde{z})$. However, we have an opposite inequality

$\lambda(\tilde{x}^0 + \varepsilon\tilde{w}) = \lambda(\tilde{x}^0) + \varepsilon|\tilde{w}|^2 > \lambda(\tilde{x}^0) = \lambda(\tilde{z})$, leading to a contradiction. In summary, we have shown that

$$\lambda(\tilde{x}) < \lambda(\tilde{z}), \forall \tilde{x} \in U. \tag{44}$$

Since $\hat{\rho}(0) = 0$, we have $0 \in U$. Letting $\tilde{x} = 0$ in (44) yields $\lambda(\tilde{z}) > 0$, so we can re-scale $\tilde{w}$ such that $\lambda(\tilde{z}) = 1 = \hat{\rho}(\tilde{z})$. Thus, (44) becomes

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

from which (36) holds. □

Lemma 4. Let $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \leq y_2 \leq \cdots \leq y_n\}$. For any fixed $\tilde{z} = (z_1, \ldots, z_n) \in \mathcal{B}$ and any $\hat{\rho}$ satisfying axioms C1, C2, E3 and C4, there exists a weight $\tilde{w} = (w_1, \ldots, w_n)$ satisfying (10), (11) and (12), such that

$$\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^{n} w_i x_i, \text{ for } \forall \tilde{x} \in \mathbb{R}^n, \text{ and } \hat{\rho}(\tilde{z}) = \sum_{i=1}^{n} w_i z_i. \tag{45}$$

Proof. The proof is obtained by using Lemma 3 and following the same line as the proof for Lemma 2. □

Proof of Theorem 4. (1) Proof for part (I). We only need to show that under condition (12), the risk statistic (13) satisfies subadditivity for any $\tilde{x}$ and $\tilde{y} \in \mathbb{R}^n$. Let $(k_1, \ldots, k_n)$ be the permutation of $(1, \ldots, n)$ such that $(\tilde{x} + \tilde{y})_{k_1} \leq (\tilde{x} + \tilde{y})_{k_2} \leq \ldots \leq (\tilde{x} + \tilde{y})_{k_n}$. Then for $i = 1, \ldots, n-1$, the partial sum up to i satisfies

$$\sum_{j=1}^{i} (\tilde{x} + \tilde{y})_{(j)} = \sum_{j=1}^{i} (\tilde{x} + \tilde{y})_{k_j} = \sum_{j=1}^{i} (x_{k_j} + y_{k_j}) \geq \sum_{j=1}^{i} (x_{(j)} + y_{(j)}). \tag{46}$$

In addition, we have for the total sum

$$\sum_{j=1}^{n} (\tilde{x} + \tilde{y})_{(j)} = \sum_{j=1}^{n} (x_j + y_j) = \sum_{j=1}^{n} (x_{(j)} + y_{(j)}). \tag{47}$$

Re-arranging the summation terms yields

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

This, along with the fact $w_i - w_{i+1} \le 0$ and equations (46) and (47), shows that

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

(2) Proof for part (II). By Axiom C4, 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 $\tilde{w} \in \mathcal{W}$ satisfying (10), (11) and (12), such that

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

where $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \le y_2 \le \cdots \le y_n\}$.

By Lemma 4, for any point $\tilde{y} \in \mathcal{B}$, there exists a weight $\tilde{w}^{(\tilde{y})}$ satisfying (10), (11) and (12) such that (45) holds. Therefore, we can take the collection of such weights as

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

Then from (45), for any fixed $\tilde{x} \in \mathcal{B}$ we have

$$\hat{\rho}(\tilde{x}) \ge \sum_{i=1}^{n} w_i^{(\tilde{y})} x_i, \ \forall \tilde{y} \in \mathcal{B}; \text{ and } \hat{\rho}(\tilde{x}) = \sum_{i=1}^{n} w_i^{(\tilde{x})} x_i.$$

Therefore,

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

which completes the proof. □

5 Proof of Theorem 5

(1) The proof for the "if" part is similar to the proof for part (I) of Theorem 1.

(2) The proof for the "only if" part: First of all, we shall prove

$$\hat{\rho}(c\tilde{x}) = c\hat{\rho}(\tilde{x}), \ \forall c \ge 0, \forall \tilde{x} \ge 0. \tag{48}$$

By Axiom F3, we have $\hat{\rho}(0) = \hat{\rho}(0) + \hat{\rho}(0)$, so

$$\hat{\rho}(0) = 0. \tag{49}$$

Axiom F3 also implies

$$\hat{\rho}(m\tilde{x}) = m\hat{\rho}(\tilde{x}),\ \forall m \in \mathbb{N}, \tilde{x} \in \mathbb{R}^n, \tag{50}$$

and $\hat{\rho}(\frac{k}{m}\tilde{x}) = \frac{1}{m}\hat{\rho}(k\tilde{x})$, $\forall m \in \mathbb{N}$, $k \in \mathbb{N} \cup \{0\}$, $\tilde{x} \in \mathbb{R}^n$, from which we have

$$\hat{\rho}(\frac{k}{m}\tilde{x}) = \frac{1}{m}\hat{\rho}(k\tilde{x}) = \frac{k}{m}\hat{\rho}(\tilde{x}),\ \forall m \in \mathbb{N}, k \in \mathbb{N} \cup \{0\}, \tilde{x} \in \mathbb{R}^n,$$

or equivalently for the set the set of nonnegative rational numbers $\mathbb{Q}^+$ we have

$$\hat{\rho}(q\tilde{x}) = q\hat{\rho}(\tilde{x}),\ \forall q \in \mathbb{Q}^+, \tilde{x} \in \mathbb{R}^n. \tag{51}$$

In general, for any $c \geq 0$ there exist two sequences $\{c_n^{(1)}\} \subset \mathbb{Q}^+$ and $\{c_n^{(2)}\} \subset \mathbb{Q}^+$, such that $c_n^{(1)} \uparrow c$ and $c_n^{(2)} \downarrow c$ as $n \to \infty$. Then for any $\tilde{x} \geq 0$, we have $c_n^{(1)}\tilde{x} \leq c\tilde{x} \leq c_n^{(2)}\tilde{x}$ for any n. It follows from Axiom F2 and (51) that

$$c_n^{(1)}\hat{\rho}(\tilde{x}) = \hat{\rho}(c_n^{(1)}\tilde{x}) \leq \hat{\rho}(c\tilde{x}) \leq \hat{\rho}(c_n^{(2)}\tilde{x}) = c_n^{(2)}\hat{\rho}(\tilde{x}),\ \forall n,\ \forall \tilde{x} \geq 0.$$

Letting $n \to \infty$, we obtain (48).

Secondly we shall show

$$\hat{\rho}(c\mathbf{1}) = c,\ \forall c \in \mathbb{R}. \tag{52}$$

By (50) and Axiom F4, we have

$$\hat{\rho}(m\mathbf{1}) = m\hat{\rho}(\mathbf{1}) = m,\ \forall m \in \mathbb{N}. \tag{53}$$

By Axiom F3, (49) and (53), we have

$$0 = \hat{\rho}(0) = \hat{\rho}(m\mathbf{1}) + \hat{\rho}(-m\mathbf{1}) = m + \hat{\rho}(-m\mathbf{1}),\ \forall m \in \mathbb{N},$$

hence

$$\hat{\rho}(-m\mathbf{1}) = -m,\ \forall m \in \mathbb{N}. \tag{54}$$

By (50),

$$\hat{\rho}(k\mathbf{1}) = \hat{\rho}(m \cdot \frac{k}{m}\mathbf{1}) = m\hat{\rho}(\frac{k}{m}\mathbf{1}),\ \forall m \in \mathbb{N}, k \in \mathbb{Z},$$

which in combination with (53) and (54) leads to

$$\hat{\rho}(\frac{k}{m}\mathbf{1}) = \frac{k}{m},\ \forall m \in \mathbb{N}, k \in \mathbb{Z}. \tag{55}$$

In general, for any $c \in \mathbb{R}$ there exist two sequences $\{c_n^{(1)}\} \subset \mathbb{Q}$ and $\{c_n^{(2)}\} \subset \mathbb{Q}$, such that $c_n^{(1)} \uparrow c$ and $c_n^{(2)} \downarrow c$ as $n \to \infty$. By Axiom F2, we have

$$\hat{\rho}(c_n^{(1)}\mathbf{1}) \leq \hat{\rho}(c\mathbf{1}) \leq \hat{\rho}(c_n^{(2)}\mathbf{1}),\ \forall n.$$

Letting $n \to \infty$ and using (55), we obtain (52).

Now we are ready to prove the theorem. Let $\tilde{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0)$ be the k-th standard basis of $\mathbb{R}^n$, $k = 1, \ldots, n$. For any $\tilde{x} \in \mathbb{R}^n$, let $(x_{(1)}, ..., x_{(n)})$ be its order statistic with $x_{(n)}$ being the largest. Then by Axiom F1 and F3, we have

$$\begin{aligned} \hat{\rho}(\tilde{x}) &= \hat{\rho}((x_{(1)}, ..., x_{(n)})) \\ &= \hat{\rho}(x_{(1)}\mathbf{1} + (0, x_{(2)} - x_{(1)}, ..., x_{(n)} - x_{(1)})) \\ &= \hat{\rho}(x_{(1)}\mathbf{1}) + \hat{\rho}((0, x_{(2)} - x_{(1)}, ..., x_{(n)} - x_{(1)})). \end{aligned}$$

Using (52), we further have

$$\begin{aligned} \hat{\rho}(\tilde{x}) &= x_{(1)} + \hat{\rho}((0, x_{(2)} - x_{(1)}, ..., x_{(n)} - x_{(1)})) \\ &= x_{(1)} + \hat{\rho}((x_{(2)} - x_{(1)}) \cdot (0, 1, \ldots, 1) + (0, 0, x_{(3)} - x_{(2)}, ..., x_{(n)} - x_{(2)})) \\ &= x_{(1)} + \hat{\rho}((x_{(2)} - x_{(1)}) \cdot (0, 1, \ldots, 1)) + \hat{\rho}((0, 0, x_{(3)} - x_{(2)}, ..., x_{(n)} - x_{(2)})), \end{aligned}$$

where the second equality follows by the permutation invariance and the last equality by the comonotonic additivity. Therefore, by (48)

$$\begin{aligned} \hat{\rho}(\tilde{x}) &= x_{(1)} + (x_{(2)} - x_{(1)})\hat{\rho}((0, 1, \ldots, 1)) + \hat{\rho}((0, 0, x_{(3)} - x_{(2)}, ..., x_{(n)} - x_{(2)})) \\ &= \cdots \\ &= x_{(1)} + (x_{(2)} - x_{(1)})\hat{\rho}((0, 1, \ldots, 1)) + (x_{(3)} - x_{(2)})\hat{\rho}((0, 0, 1, \ldots, 1)) + \cdots \\ &\quad + (x_{(n)} - x_{(n-1)})\hat{\rho}((0, ..., 0, 1)) \\ &= \sum_{i=1}^{n} w_i x_{(i)}, \end{aligned}$$

where

$$w_i = \hat{\rho}(\sum_{j=i}^{n} \tilde{e}_j) - \hat{\rho}(\sum_{j=i+1}^{n} \tilde{e}_j), i = 1, \ldots, n,$$

with $\tilde{e}_j$ being a vector such that the jth element is one and all other elements are zero. Since $\sum_{i=1}^{n} w_i = \hat{\rho}(\sum_{j=1}^{n} \tilde{e}_j) = 1$ and $w_i \geq 0, i = 1, \ldots, n$, by Axiom F2, the proof is completed. □

6 Proof of Proposition 1

The result of TCM is from equation (3.2.3) in [55]. To show (17), note that by equation (3.2.4) in [55] the influence function of the $(1 - \alpha)$-trimmed mean $T_{1-\alpha}(X) := E[X|X <$

$\mathrm{VaR}_\alpha(X)]$ is

$$IF(x, T_{1-\alpha}, X) = \begin{cases} \frac{x-(1-\alpha)\mathrm{VaR}_\alpha(X)}{\alpha} - E[X|X < \mathrm{VaR}_\alpha(X)], & \text{if } x \leq \mathrm{VaR}_\alpha(X) \\ \mathrm{VaR}_\alpha(X) - E[X|X < \mathrm{VaR}_\alpha(X)] & \text{if } x > \mathrm{VaR}_\alpha(X) \end{cases} \tag{56}$$

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

$$IF(x, E[X], X) = x - E[X]. \tag{57}$$

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

$$IF(x, E[X], X) = \alpha IF(x, T_{1-\alpha}, X) + (1-\alpha) IF(x, \mathrm{TCE}_\alpha, X). \tag{58}$$

Now (17) follows from equations (56), (57) and (58). □