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

Xianhua Peng
The Hong Kong University of Science and Technology

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

C. C. Heyde, Steven Kou†, Xianhua Peng
Columbia University

First draft July 2006; this version February 2009

Abstract

Choosing a proper risk measure is of great regulatory importance, as exemplified in Basel Accord that uses Value-at-Risk (VaR) in combination with scenario analysis as a preferred risk measure. The main motivation of this paper is to investigate whether VaR, in combination with scenario analysis, is a good risk measure for external regulation. While many risk measures may be suitable for internal management, we argue that risk measures used for external regulation should have robustness with respect to modeling assumptions and data. We propose new data-based risk measures called natural risk statistics that are characterized by a new set of axioms based on the comonotonicity from decision theory. Natural risk statistics include VaR as a special case and therefore provide a theoretical basis for using VaR along with scenario analysis as a robust risk measure for the purpose of external, regulatory risk measurement.

Keywords: risk measures, decision theory, prospect theory, tail conditional expectation, tail conditional median, value at risk, quantile, robust statistics.

JEL classification: G18, G28, G32, K20, K23.

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

†Corresponding author. Address: 312 S. W. Mudd Building, 500 West 120th Street, New York, NY 10027, USA. Phone: +1 212-854-4334. Fax: +1 212-854-8103. Email: sk75@columbia.edu.

1 Introduction

Choosing a proper risk measure is an important regulatory issue, as exemplified in governmental regulations such as Basel II (Basel Committee, 2006), which uses VaR (or quantiles) along with scenario analysis as a preferred risk measure. The main motivation of this paper is to investigate whether VaR, in combination with scenario analysis, is a good risk measure for external regulation. By using the notion of comonotonic random variables studied in decision theory literatures such as Schmeidler (1986, 1989), Yaari (1987), Tversky and Kahneman (1992), Wakker and Tversky (1993), and Wang, Young, and Panjer (1997), we shall postulate a new set of axioms that are more general and suitable for external, regulatory risk measures. In particular, 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 to use 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 other 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.

In this paper we only study static risk measures, i.e., one period risk measures. Mathematically, let Ω be the set of all possible states of nature at the end of an observation period, $\mathcal{X}$ be the set of financial losses under consideration, in which each financial loss is a random variable defined on Ω. Then a risk measure ρ is a mapping from $\mathcal{X}$ to the real line $\mathbb{R}$. The multi-period or dynamic risk measures are related to dynamic consistency or time-consistency for preferences; see for instance, Koopmans (1960); Epstein and Zin (1989); Duffie and Epstein (1992); Wang (2003); Epstein and Schneider (2003).

There are two main families of static risk measures suggested in the literature, coherent risk measures proposed by Artzner, Delbaen, Eber, and Heath (1999) and insurance risk measures proposed by Wang, Young, and Panjer (1997). To get a coherent risk measure, one first chooses a set of scenarios (different probability measures), and then defines the coherent risk measure as the maximal expected loss under these scenarios. To get an insurance risk measure, one first fixes a distorted probability, and then defines the insurance risk measure as the expected loss under the distorted probability (only one scenario). Both approaches are axiomatic, in the sense that some axioms are postulated first, and then all the risk measures satisfying the axioms are 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.

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

The objective of a risk measure is an important issue that has not been well addressed in the existing literature. In terms of objectives, risk measures can be classified into two categories: internal risk measures used for firms' internal risk management, and external risk measures used for governmental regulation. Internal risk measures are proposed for the interests of a firm's managers or/and equity shareholders, while external risk measures are used by regulatory agencies to maintain safety and soundness of the financial system. One risk measure may be suitable for internal management, but not for external regulation, and vice versa. There is no reason to believe that one unique risk measure can fit all needs.

In this paper, we shall focus on external risk measures from the viewpoint of governmental/regulatory agencies. We will show in Sections 3 and Section 4 that an external risk measure should be robust, i.e., it should not be very sensitive to modeling assumptions or to outliers in the data.[1] In addition, external risk measures should emphasize the use of data (historical and simulated) rather than solely depend on subjective models.

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 new risk measures that

[1]Mathematically, robustness has been studied extensively in the statistics literature, e.g. Huber (1981) and Staudte and Sheather (1990).

satisfy the new set of axioms are fully characterized in the paper. More precisely, the contribution of this paper is eightfold.

(1) We give the reasons why a different set of axioms are needed: (a) We present some critiques of subadditivity from the viewpoints of law and robustness (see Section 3 and Section 4), as well as from the viewpoints of diversification, bankruptcy protection, and decision theory (see Section 5). (b) The main drawback of insurance risk measures is that they do not incorporate scenario analysis; i.e., unlike coherent risk measures, an insurance risk measure chooses a single 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 (historical and simulated). Our approach is based on data rather than on some hypothetical distributions.

(2) A new set of axioms based on data and comonotonic subadditivity are postulated and natural risk statistics are defined in Section 6. A complete characterization of natural risk statistics is given in Theorem 1.

(3) Coherent risk measures rule out the use of VaR. Theorem 1 shows that natural risk statistics include VaR as a special case and thus give an axiomatic justification to the use of VaR in combination with scenario analysis.

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

(5) Theorem 3 and Theorem 4 in Section 7.1 completely characterize data-based coherent risk measures and data-based law-invariant coherent risk measures. Theorem 4 shows that coherent risk measures are in general not robust.

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

(7) We point out in Section 4 that natural risk statistics include tail conditional median as a special case, which is more robust than tail conditional expectation suggested by coherent risk measures.

(8) In the existing literature, some examples are used to show that VaR does not satisfy subadditivity. However, we show in Section 8 if one considers the tail conditional median, or equivalently considers VaR at a higher level, the problem of non-subadditivity of VaR disappears.

2 Review of Existing Risk Measures

2.1 Value-at-Risk

One of the most widely used risk measures in financial regulation and risk management is Value-at-Risk (VaR), which is a quantile at some pre-defined probability level. More precisely, given $\alpha \in (0,1)$, VaR of the loss variable X at level α is defined as the α-quantile of X, i.e.,

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

The banking regulation Basel II (Basel Committee, 2006) specifies VaR at 99th percentile as a preferred risk measure for measuring market risk.

2.2 Coherent and Convex Risk Measures

2.2.1 Subadditivity and Law Invariance

Artzner, Delbaen, Eber, and Heath (1999) propose the coherent risk measures that satisfy the following three axioms:

Axiom A1. Translation invariance and positive homogeneity:

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

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

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

Axiom A1 states that the risk of a financial position is proportional to the size of the position, and a sure loss of amount b simply increases the risk by b. This axiom is proposed from the accounting viewpoint. For many external risk measures, such as margin deposit and capital requirement, the accounting-based axiom seems to be reasonable. Axiom A1 is relaxed in convex risk measures,[2] by which the risk of a financial position does not necessarily increases linearly with the size of that position. Axiom A2 is a minimum requirement for a reasonable risk measure.

What is questionable lies in the subadditivity requirement in Axiom A3, which basically means that "a merger does not create extra risk" (Artzner, Delbaen, Eber, and Heath, 1999, pp. 209). We will discuss the controversies related to this axiom in Section 5.

[2] Convex risk measures are proposed by Föllmer and Schied (2002) and independently by Frittelli and Gianin (2002), in which 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 $X, Y \in \mathcal{X}$, $\lambda \in [0,1]$.

Artzner, Delbaen, Eber, and Heath (1999) point out that it is shown by Huber (1981) 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 (2002) extends the above result to general probability spaces with infinite number of states, assuming the Fatou property. Therefore, measuring risk by a coherent risk measure amounts to computing maximal expectation under different scenarios (different Qs), thus justifying scenario analysis used in practice. Artzner, Delbaen, Eber, and Heath (1999) and Delbaen (2002) also present an equivalent approach of defining coherent risk measures via acceptance sets.

A special case of coherent risk measures is the tail conditional expectation (TCE).[3] More precisely, the TCE of the loss X at level α is defined by

$$\mathrm{TCE}_\alpha(X) = \text{mean of the } \alpha\text{-tail distribution of } X. \tag{2}$$

If the distribution of X is continuous, then

$$\mathrm{TCE}_\alpha(X) = E[X | X \geq \mathrm{VaR}_\alpha(X)]. \tag{3}$$

For discrete distributions, $\mathrm{TCE}_\alpha(X)$ is a regularized version of $E[X|X \geq \mathrm{VaR}_\alpha(X)]$. TCE satisfies subadditivity for continuous random variables, and also for discrete random variables if one defines quantiles for discrete random variables properly (see Acerbi and Tasche, 2002).

Another 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.

Law invariance means that the risk of a position is determined purely by the loss distribution of the position. A risk measure is called a law-invariant coherent risk measure, if it satisfies Axiom A1-A4. Kusuoka (2001) gives a representation for the law-invariant coherent risk measures.

[3] Tail conditional expectation is also called expected shortfall by Acerbi, Nordio, and Sirtori (2001) and conditional value-at-risk by Pflug (2000) and Rockafellar and Uryasev (2000), respectively.

2.2.2 Main Drawbacks

A main drawback of coherent risk measures is that in general they are not robust (see Section 7.1). For example, TCE is too sensitive to the modeling assumption for the tails of loss distributions, and it is sensitive to outliers in the data (see Section 4.3). However, robustness is an indispensable requirement for external risk measures, as we will point out in Section 4. Hence, coherent risk measures may not be suitable for external regulation.

Another drawback of coherent risk measures and convex risk measures is that they rule out the use of VaR as risk measures, because VaR may not satisfy subadditivity (see Artzner, Delbaen, Eber, and Heath, 1999). 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 inconsistency is due to the subadditivity Axiom A3. By relaxing this axiom and requiring subadditivity only for comonotonic random variables, we are able to find a new set of axioms and define the new risk measures that include VaR and quantiles, thus eliminating the inconsistency.

2.3 Insurance Risk Measures

Wang, Young, and Panjer (1997) propose the insurance risk measures that satisfy the following five axioms:

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

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

Axiom B3. Comonotonic additivity: $\rho(X+Y) = \rho(X) + \rho(Y)$, if X and Y are comonotonic. (X and Y are comonotonic if $(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0$ 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 $x^+ \triangleq \max(x, 0)$, $\forall x \in \mathbb{R}$.

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

The notion of comonotonic random variables is studied in Schmeidler (1986), Yaari (1987) and Denneberg (1994). 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. The psychological motivation of comonotonic random variables comes from alternative models to expected utility theory including the prospect theory (see Section 5.3). Dhaene, Vanduffel, Goovaerts, Kaas, Tang, and Vyncke (2006) give a recent review of risk measures and comonotonicity.

Wang, Young, and Panjer (1997) prove that if $\mathcal{X}$ contains all the Bernoulli(p) random variables, $0 \leq p \leq 1$, then the risk measure ρ satisfies Axiom 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 that is nondecreasing and satisfies $g(0) = 0$ and $g(1) = 1$, and $g \circ P$ is called the distorted probability that is defined by $g \circ P(A) \triangleq g(P(A))$, for any event A. The detailed discussion of Choquet integration can be found in Denneberg (1994).

It should be emphasized that VaR satisfies Axioms B1-B5 (see Corollary 4.6 in Denneberg, 1994, for the proof that VaR satisfies Axiom B4) and hence is an insurance risk measure. In general, an insurance risk measure in (4) does not satisfy subadditivity, unless the distortion function $g(\cdot)$ is concave (see Denneberg, 1994).

A main drawback of insurance risk measures is that they do not incorporate scenario analysis. Unlike coherent risk measures, an insurance risk measure corresponds to a fixed distortion function g and a fixed probability measure P, so it does not allow one to compare different distortion functions or different probability measures.

The main reason that insurance risk measures rule out scenario analysis is that they require comonotonic additivity. Wang, Young, and Panjer (1997) impose comonotonic additivity based on the argument that comonotonic random variables do not hedge against each other. However, this is true only if one focuses on one scenario. The counterexample at the end of Section 7 shows that if one incorporates different scenarios, one can get strict subadditivity rather than additivity for comonotonic random variables. Hence, Axiom B3 may be too restrictive. To incorporate scenario analysis, we shall relax the comonotonic additivity to comonotonic subadditivity.

The mathematical concept of comonotonic subadditivity is also studied independently by Song and Yan (2006a), who give a representation of the functionals satisfying comonotonic subadditivity or comonotonic convexity from a mathematical perspective. Song and Yan (2006b) give a representation of risk measures that respect stochastic orders and are comonotonically subadditive or convex.

There are several differences between Song and Yan (2006a,b) and this paper. First, we provide a full mathematical characterization of the new risk measures that are based on data rather than on some hypothetical distributions. Second, we provide alternative axioms to characterize the new risk measures through acceptance sets. Third, we provide legal, economic, and psychological reasons for postulating the comonotonic subadditivity axiom. Fourth, 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. Fifth, we give some counterexamples in Section 8 showing that the tail conditional median can satisfy subadditivity when VaR fails to do so.

3 Legal Motivation

In this section, we shall summarize some basic concepts of law that motivate us to propose a new set of axioms for external risk measures. Since axioms are always subject to debate and change, proposing a new set of axioms is useful in that it provides people with more choices to apply in different 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 external risk measures. By understanding basic concepts of law, we will have a better understanding of desirable properties of external risk measures. We shall point out: (1) External risk measures should be robust, because robustness is essential for law enforcement. (2) External risk measures 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, 1994). Two concepts to be discussed here are legal realism and legal positivism. The former concerns the robustness of law, while the latter concerns the consistency of law to social norms.

3.1 Legal Realism and Robustness of Law

Legal realism is the viewpoint that the legal decision of a court regrading a case is determined by the actual practices of judges, rather than the law set forth in statutes and precedents. All legal rules contained in statutes and precedents have uncertainty due to the uncertainty in human language and that human beings are unable to anticipate all possible future circumstances (Hart, 1994, pp. 128). Hence, a law is only a guideline for judges and enforcement officers (Hart, 1994, pp. 204–205), i.e., a law is only intended to be the average of what judges and officers will decide. This concerns the robustness of law, i.e., a law should be established in such a way that different judges will reach similar conclusions when they implement the law.

In particular, the enforcement of a risk measure in banking regulation requires that the risk measure should be robust with respect to underlying models and data; otherwise, it is very hard for the government to enforce the risk measure upon firms. Furthermore, without robustness, the risk measure may even be vulnerable to firms' manipulation; for example, a firm may choose a model that significantly reduces the regulatory capital requirement. However, coherent risk measures generally lack robustness, as discussed in Section 4 and manifested in Theorem 4 (see Section 7.1). Section 4.3 shows that tail condition median, a special case of the proposed new risk measures, is more robust than 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. It is based on the argument that if a system of rules are to be imposed by force in 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 (Hart, 1994, pp. 201–204).

Therefore, risk measures imposed in banking regulations should also reflect most people's behavior; otherwise, the regulation cannot be enforced. However, decision theory suggests that most people's decision can violate the subadditivity Axiom A3 (see Section 5.3 for details). This motivates us to propose the new risk measures that are consistent with most people's behavior.

3.3 An Example of Speed Limit

An illuminating example manifesting the above ideas is setting up speed limit on the road, which is a crucial issue involving life and death decisions. In 1974, the U.S. Congress enacted the 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 speed limits on certain roads were increased to 65 mph in 1987. In 1995, the law was repealed and the choice of speed limit was returned to each state, in part because of notoriously low compliance.

Today, the "Manual on Uniform Traffic Control Devices" of American Association of State Highway and Transportation Officials recommends setting speed limit near the 85th percentile speed of free flowing traffic with an adjustment taking into consideration that people tend to drive 5 to 10 miles above the posted speed limit.[4] This recommendation is adopted by all states and most local agencies.[5] 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) The laws that reflect the behaviors of the majority of drivers are found to be successful, while the laws that lack the consent and voluntary compliance of the public majority cannot be effectively enforced.

4 The Main Reason to Relax Subadditivity: Robustness

4.1 Robustness is Indispensable for External Risk Measures

In determining capital requirement, regulators allow firms to use their own internal risk models and private data in the calculation of regulatory capital. The reliance on the firms' internal models and private data is largely due to the fact that only the firms themselves have the necessary information to identify and evaluate all relevant risks. For example, Basel II's advanced approaches allow firms to use their own internal models to calculate the capital requirement for the market risk.

There are two issues rising from the use of internal models and private data in external regulation: (1) The data can be noisy, flawed or unreliable. (2) There can be several feasible models for the same asset or portfolio, as the true model may never be identified from the data.

[4] See Transportation Research Board of the National Academies (2003), pp. 51.

[5] See Institute of Transportation Engineers (2000).

For example, the heaviness of tail distributions cannot be identified in many cases. Although it is accepted that stock returns have tails heavier than those of normal distribution, one school of thought believes that the tails are exponential-type, while another believes that the tails are power-type. Heyde and Kou (2004) show 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 a larger 99th 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 that 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.

To address the two issues above, external risk measures should demonstrate robustness with respect to underlying models and data. When an external risk measure is imposed by a regulator, it must be unambiguous, stable, and be able to be implemented consistently throughout all the relevant firms, no matter what internal beliefs or internal models each individual firm may have. In situations when the correct model cannot be identified, two firms that have exactly the same portfolio can use different internal models, both of which can obtain the approval of the regulator. From the regulator's point of view, the same portfolio should incur the same or at least almost the same amount of regulatory capital. Therefore, the external risk measure should be robust, otherwise different firms can be required to hold very different regulatory capital for the same risk exposure, which makes the risk measure unacceptable to most firms. In addition, if the external risk measure is not robust, firms can take regulatory arbitrage by choosing a model that significantly reduces the regulatory capital requirement.

4.2 Coherent Risk Measures Are Not Robust

The robustness of coherent risk measures is questionable in two aspects.

First, the theory of coherent risk measures suggests to use TCE to measure risk. However, as we will show in Section 4.3, TCE is sensitive to model assumptions of heaviness of tail distribution and to outliers in the data.

Second, in general, coherent risk measures are not robust. 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. We will prove in Theorem 4 (in Section 7) that any empirically law-invariant coherent risk measure $\hat{\rho}(\tilde{x})$ can be represented by

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

where $\mathcal{W} = \{\tilde{w}\}$ is a set of nonnegative weights with each $\tilde{w} = (w_1, \ldots, w_n) \in \mathcal{W}$ satisfying $w_1 \leq w_2 \leq \cdots \leq w_n$. Because the risk measure in formula (5) puts larger weights on larger observations, it is obviously not robust. Therefore, coherent risk measures are generally not robust. An extreme case of coherent risk measures is the maximum loss $x_{(n)}$, which is not robust at all.

4.3 Tail Conditional Median: a Robust Risk Measure

We propose a more robust risk measure, tail conditional median (TCM), for external regulation. TCM of the loss X at level α is defined as

$$\mathrm{TCM}_\alpha(X) = \mathrm{median}[X | X \geq \mathrm{VaR}_\alpha(X)]. \tag{6}$$

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

TCM is a special case of the new risk measures to be defined in Section 6. If X has a continuous distribution, then

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

This shows that VaR at a higher level can incorporate tail information, which contradicts some claims in the existing literature. For example, if one wants to measure the size of losses beyond 95% level, one can use VaR at 97.5%, which is TCM at 95% level. For discrete random variables or data, one simply uses (6) to define TCM and there may be a difference between $\mathrm{TCM}_\alpha(X)$ and $\mathrm{VaR}_{\frac{1+\alpha}{2}}(X)$, depending on how quantiles of discrete data are defined.

There are many examples in the existing literature showing that VaR violates subadditivity. However, Section 8 shows that in those examples subadditivity will not be violated if one replaces VaR by TCM.

Tail conditional median is more robust than tail conditional expectation, as demonstrated in Figure 1. The left panel of the figure shows the value of TCE_α with respective to $\log(1-\alpha)$ for Laplacian distribution and T distribution, where α is in the range $[0.95, 0.999]$. The right panel of the figure shows the value of TCM_α with respective to $\log(1-\alpha)$ for Laplacian 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 underlying distribution. For example, with $\alpha = 99.6\%$, the variation of TCE_α with respect to the change of underlying models is 1.44, whereas the variation of TCM_α is only 0.75.

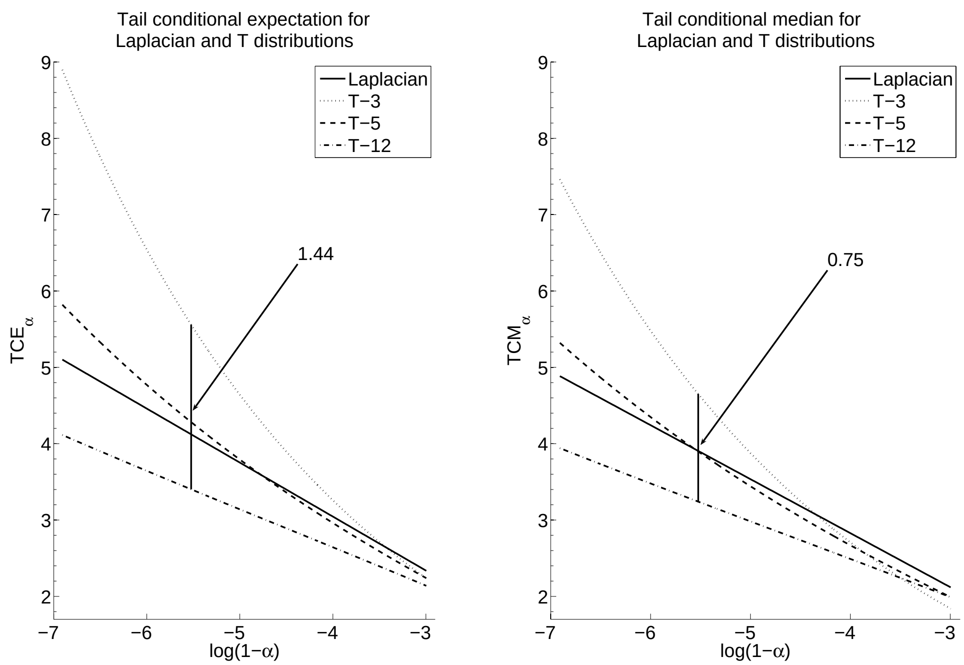

Figure 1: Comparing the robustness of tail conditional expectation and tail conditional median. The distributions used are Laplacian distribution and T distributions with degree of freedom 3, 5, 12, normalized to have mean 0 and variance 1. The x-axis is $\log(1-\alpha)$ where $\alpha \in [0.95, 0.999]$.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

The robustness of TCM can also be theoretically justified by its bounded influence function. The influence function is introduced by Hampel (1974) as a useful tool for 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., $\sup_x IF(x, T, F) < \infty$. If the influence function of an estimator $T(F)$ is unbounded, then an outlier in the data may cause problems. By simple calculation,[6] TCE has an unbounded influence function but TCM has a bounded influence function, which implies that TCM is more robust than TCE with respect to outliers in the data.

4.4 Robust Risk Measures vs. Conservative Risk Measures

A risk measure is said to be more conservative than another one, if it generates higher risk measurement than the other for the same underlying risk exposure. The use of more conservative risk measures in external regulation is desirable from a regulator's viewpoint, since it generally increases the safety of the financial system (Of course, risk measures should not be too conservative to retard the economic growth).

There is no contradiction between the robustness and the conservativeness of external risk measures. Robustness addresses the issue of whether a risk measure can be implemented consistently, so it is a requisite property of an external risk measure. Conservativeness addresses the issue of how heavily an external risk measure should be implemented to regulate the financial market, given that the external risk measure can be implemented consistently. In other words, being more conservative is a further desirable property of an external risk measure. An external risk measure should be robust in the first place before one can consider the issue of how to implement it in a conservative way.

For two reasons it is wrong to argue that TCE is more conservative than TCM and hence is more suitable for external regulation:

(1) It is not true that TCE is more conservative than TCM. Indeed, it is not clear which of the two risk measures gives higher risk measurement. Eling and Tibiletti (2008) compare TCE and TCM for a set of capital market data, including the returns of S&P 500 stocks, 1347 mutual funds and 205 hedge funds, and find that although

[6] See Appendix E.

TCE is on average about 10% higher than TCM at standard confidence levels, TCM is higher than TCE in about 10% of all cases. Therefore, TCE is not necessarily more conservative than TCM.

(2) Because TCE is not robust, from a legal viewpoint it cannot be implemented consistently. Therefore, TCE is not suitable for external regulation.

A feasible approach to combine robustness and conservativeness is to implement a robust risk measure in a conservative way as follows:

(1) Stress testing and scenario analysis can be used to generate data that incorporate the possible effects of extreme events that lie outside normal market conditions. The generated data can be combined with the original historical data for the calculation of risk measure.

(2) Regulators can simply multiply the original risk measure by a larger constant to obtain a more conservative risk measure. For example, in Basel II, the regulatory capital for market risk is defined as $k \cdot \text{VaR}$, where k is a constant bigger than 3. The regulator can choose a larger k in the implementation. This is equivalent to postulating $\rho(1) = k$ instead of $\rho(1) = 1$ as an axiom for defining the risk measure ρ.

(3) One can use more weights in the definition of the risk measure. More precisely, it is shown in Section 6 that a natural risk statistic can be represented as

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

where $\mathcal{W} = \{\tilde{w} = (w_1, \ldots, w_n)\}$ is a set of nonnegative weights. In contrast to a coherent risk statistic represented in equation (5), each weight $\tilde{w}$ here does not need to satisfy $w_1 \leq w_2 \leq \cdots \leq w_n$. By including more weights $\tilde{w}$ in the set $\mathcal{W}$, one can get a natural risk statistic that gives more conservative risk measurement.

5 Other Reasons to Relax Subadditivity

5.1 Diversification and Tail Subadditivity of VaR

The subadditivity axiom conforms to the idea that diversification does not increase risk.[7] 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

[7] The convexity axiom for convex risk measures also comes from the idea of diversification.

other is based on expected utility theory. Samuelson (1967) shows that any investor with a strictly concave utility function will uniformly diversify among independently and identically distributed (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 constructed from these i.i.d. risks.[8] 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 (2006) show 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. They also show that, investors with certain S-shaped utility functions would prefer non-diversification, even for bounded risks. A S-shaped utility function is convex in the domain of losses. The convexity in the domain of losses is supported by experimental results and prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992), which is an important alternative to expected utility theory. See Section 5.3 for more discussion on prospect theory.

VaR has been criticized because it does not have subadditivity universally, but this criticism is unreasonable because even diversification is not universally preferable. In fact, Daníelsson, Jorgensen, Samorodnitsky, Sarma, and de Vries (2005) show 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).[9] The simulations that they carry out also show 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. The consistency between the use of VaR and diversification is shown in Table 1.

Asset returns with tail index less than one have very fat tails. They are hard to

[8] See Brumelle (1974); McMinn (1984); Hong and Herk (1996); Kijima (1997) for the discussion on whether diversification is beneficial when the asset returns are dependent.

[9] More precisely, Daníelsson, Jorgensen, Samorodnitsky, Sarma, and de Vries (2005) prove 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 $\text{VaR}_\alpha(X+Y) \le \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y)$, $\forall \alpha \in (\alpha_0, 1)$. (2) If the tail index of the X and Y are different, then a weaker form of tail subadditivity holds: $\limsup_{\alpha \to 1} \frac{\text{VaR}_\alpha(X+Y)}{\text{VaR}_\alpha(X)+\text{VaR}_\alpha(Y)} \le 1$.

	Not very fat tails	Fat tails
Does diversification help to reduce risk?	Yes	No
Does VaR satisfy subadditivity?	Yes	No

Table 1: VaR is consistent with diversification.

find and easy to identify. Daníelsson, Jorgensen, Samorodnitsky, Sarma, and de Vries (2005) argue 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, Renault, and Tsafack (2007) show 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 A Merger Always Reduce Risk

Subadditivity basically means that "a merger does not create extra risk" (Artzner, Delbaen, Eber, and Heath, 1999, pp. 209). However, Dhaene, Goovaerts, and Kaas (2003) point out that a merger may increase risk, particularly due to bankruptcy protection for firms. For example, it is better to split a risky trading business into a separate subsidiary firm. This way, even if the loss from the subsidiary firm is enormous, the parent firm can simply let the subsidiary firm go bankrupt, thus confining the loss to that one subsidiary firm. Therefore, creating subsidiary firms may incur less risk and a merger may increase risk.[10]

For example, the collapse of Britain's Barings Bank (which had 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 a 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.

[10]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 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, given that everything else is equal, a merger increases the risk of investment. This contradicts the intuition that "A merger does not create extra risk".

In addition, there is little empirical evidence supporting the argument that "a merger does not create extra risk". Indeed, in practice, the credit rating agencies, such as Moody's and Standard & Poor's, do not upgrade a firm's credit rating because of a merger. On the contrary, the credit rating of the joint firm may be lowered shortly after the merger of two firms; the recent merger of Bank of America and Merrill Lynch is such an example.

5.3 Psychological Theory of Uncertainty and Risk

Risk measures have a close connection with the psychological theory of people's preferences of uncertainties and risks. Kahneman and Tversky (1979) and Tversky and Kahneman (1992) point out that people's choices under risky prospects are inconsistent with the basic tenets of expected utility theory and propose an alternative model, called prospect theory, which can explain a variety of preference anomalies including the Allais and Ellsberg paradoxes. The axiomatic analysis of prospect theory is presented in Tversky and Kahneman (1992) and extended in Wakker and Tversky (1993). Many other people have studied alternative models to expected utility theory, such as Quiggin (1982), Schmeidler (1986), Yaari (1987), Schmeidler (1989), and so on. These models are referred to as "anticipated utility theory", "rank-dependent models", "cumulative utility theory", etc.

Prospect theory implies fourfold pattern of risk attitudes: people are risk averse for gains with high probability and losses with low probability; and people are risk seeking for gains with low probability and losses with high probability. Prospect theory postulates that: (1) The value function (utility function) is normally concave for gains, convex for losses, and is steeper for losses than for gains. (2) People evaluate uncertain prospects using "decision weights" that are nonlinear transformation of probabilities and can be viewed as distorted probabilities.

Schmeidler (1989) points out that risk preferences for *comonotonic* random variables are easier to justify than those for arbitrary random variables and accordingly relaxes the independence axiom in expected utility theory to the comonotonic independence axiom that applies only to comonotonic random variables. Tversky and Kahneman (1992) and Wakker and Tversky (1993) also find that prospect theory is characterized by the axioms that impose conditions only on comonotonic random variables and are thus less restrictive than their counterparts in expected utility theory.

There are simple examples showing that risk associated with non-comonotonic random variables can violate subadditivity, because people are risk seeking instead of risk averse in choices between probable and sure losses, as implied by prospect theory.[11] [12]

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

6 Main Results: Natural Risk Statistics and Their Axiomatic Representations

Suppose we have a collection of data $\tilde{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ related to a random loss X, which can be discrete or continuous. $\tilde{x}$ can be a set of historical observations, or a set of simulated data generated according to a well-defined procedure or model, or a mixture of historical and simulated data. To measure risk from the data $\tilde{x}$, we define a risk statistic $\hat{\rho}$ to be a mapping from $\mathbb{R}^n$ to $\mathbb{R}$. A risk statistic is a data-based

[11] Suppose there is an urn that contains 50 black balls and 50 red balls. Randomly draw a ball from the urn. Let B be the position of losing \$10,000 in the event that the ball is black, and R be the position of losing \$10,000 in the event that the ball is red. Obviously, B and R bear the same amount 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 prospect theory, people are risk seeking in choices between probable and sure losses, i.e., most people would prefer a larger loss with a substantial probability to a sure loss. Therefore, most people would prefer position B to position S (see problem 12 on pp. 273 in Kahneman and Tversky (1979), and table 3 on pp. 307 in Tversky and Kahneman (1992)). 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 violate subadditivity.

[12] Even in terms of expected utility theory, it is not clear whether a risk measure should be superadditive or subadditive for independent random variables. For instance, Eeckhoudt and Schlesinger (2006) link the sign of utility functions to risk preferences. Let $u^{(4)}$ be the fourth derivative of an utility function u. They prove that $u^{(4)} \leq 0$ if and only if $E(u(x+\epsilon_1+\epsilon_2)) + E(u(x)) \leq E(u(x+\epsilon_1)) + E(u(x+\epsilon_2))$, for any $x \in \mathbb{R}$ and any independent risks ϵ_1 and ϵ_2 satisfying $E(\epsilon_1) = E(\epsilon_2) = 0$. This result can be interpreted as follows. Suppose the owner of two subsidiary firms, each of which has initial wealth x, faces the problem of assigning two projects to the two firms. The net payoffs 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 firm or prefers to assign one project to each firm depends on the sign of the fourth derivative of his utility function. Because comonotonic random variables are generally not independent, we shall impose subadditivity only for comonotonic random variables to avoid the issue of possible superadditivity for independent random variables.

risk measure.

In this section, we will define and fully characterize the new data-based risk measure, which we call the natural risk statistic, through two sets of axioms and two representations.

6.1 The First Representation

At first, we postulate the following axioms for a risk statistic $\hat{\rho}$.

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 $x_i\leq y_i, i=1,\dots,n$.

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

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

and Axioms C1 and C2 imply $\hat{\rho}$ is continuous.[13]

Axiom C3. Comonotonic subadditivity:

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

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 C3 we relax the subadditivity requirement in coherent risk measures so that the axiom is only enforced for comonotonic random variables. This also relaxes the comonotonic additivity requirement in insurance risk measures. Comonotonic subadditivity is consistent with prospect theory, as we explained in Section 5.3.

Axiom C4. Permutation invariance:

$$\hat{\rho}((x_1,\dots,x_n))=\hat{\rho}((x_{i_1},\dots,x_{i_n}))$$

for any permutation $(i_1,\dots,i_n)$ of $(1,2,\dots,n)$.

[13]Indeed, suppose $\hat{\rho}$ satisfies Axiom C1 and C2. Then for any $\tilde{x}\in\mathbb{R}^n$, $\varepsilon>0$, and $\tilde{y}$ satisfying $|y_i-x_i|<\varepsilon, i=1,\dots,n$, 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 $\hat{\rho}$.

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 measurement of risk.

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

The following representation theorem for natural risk statistics is one of the main results of this paper.

Theorem 1 *Let $x_{(1)}, ..., x_{(n)}$ be the order statistics of the data $\tilde{x}$ 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

$$\hat{\rho}(\tilde{x}) \triangleq \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^{n} w_i x_{(i)}\}, \ \forall \tilde{x} \in \mathbb{R}^n \tag{7}$$

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

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^{n} w_i x_{(i)}\}, \ \forall \tilde{x} \in \mathbb{R}^n. \tag{8}$$

Proof. See Appendix A. *Q.E.D.*

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 $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \leq y_2 \leq \cdots \leq y_n\}$. However, unlike in the case of coherent risk measures, the existence of a set of weights $\mathcal{W}$ such that (8) holds does not follow easily from the proof in Huber (1981). The main difference here is that the comonotonic set $\mathcal{B}$ is not an open set in $\mathbb{R}^n$. The boundary points do not have as nice properties as the interior points. 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 $w_i \geq 0$ for $i = 1, \ldots, n$.[14]

[14] Utilizing convex duality theory, Ahmed, Filipović, and Svindland (2008) provide alternative shorter proofs for Theorem 1 and Theorem 4 after seeing the first version of this paper.

Each weight $\tilde{w}$ in the set $\mathcal{W}$ in equation (7) represents a scenario, so natural risk statistics can incorporate scenario analysis by putting different weights on the order statistics. In addition, since estimating VaR from data is equivalent to calculating weighted average of the order statistics, Theorem 1 shows that natural risk statistics include VaR along with scenario analysis as a special case.

6.2 The Second Representation via Acceptance Sets

The natural risk statistics can also be characterized via acceptance sets as in the case of coherent risk measures. More precisely, a statistical acceptance set is a subset of $\mathbb{R}^n$ that includes all the data that are considered acceptable by a regulator. 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 in order for the resulting position to be acceptable, or in mathematical form

$$\hat{\rho}_{\mathcal{A}}(\tilde{x}) = \inf\{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{A}\},\ \forall \tilde{x} \in \mathbb{R}^n. \tag{9}$$

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

$$\mathcal{A}_{\hat{\rho}} = \{\tilde{x} \in \mathbb{R}^n \mid \hat{\rho}(\tilde{x}) \le 0\}. \tag{10}$$

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 statistical 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 statistical 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 statistical 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 $(i_1, \ldots, i_n)$ of $(1, 2, \ldots, n)$.

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

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

Proof. See Appendix B. *Q.E.D.*

7 Comparison with Coherent 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 define the coherent risk statistics, the data-based versions of coherent risk measures.

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

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

We have the following representation theorem for coherent risk statistics.

Theorem 3 *A risk statistic $\hat{\rho}$ 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 $w_i \geq 0, i = 1, \ldots, n$, such that*

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^n w_i x_i\}, \ \forall \tilde{x} \in \mathbb{R}^n. \tag{11}$$

Proof. The proof for the "if" part is trivial. To prove 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 $\mathcal{Z}$: $E^*(Z) \triangleq \hat{\rho}(Z(\theta_1), Z(\theta_2), \ldots, Z(\theta_n))$. 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 (1981). Then the result follows by applying Proposition 2.1 at page 254 of Huber (1981) to E^*. *Q.E.D.*

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 are the counterparts of law-invariant coherent risk measures.

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

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

Theorem 4 *Let $x_{(1)}, ..., x_{(n)}$ be the order statistics of the data $\tilde{x}$ 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

$$\sum_{i=1}^{n} w_i = 1, \tag{12}$$

$$w_i \geq 0, i = 1, \ldots, n, \tag{13}$$

$$w_1 \leq w_2 \leq \ldots \leq w_n, \tag{14}$$

the risk statistic

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

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 (12), (13) *and* (14)*, such that*

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^{n} w_i x_{(i)}\}, \ \forall \tilde{x} \in \mathbb{R}^n. \tag{16}$$

Proof. See Appendix C. *Q.E.D.*

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

(1) A natural risk statistic is the supremum of a set of L-statistics (a L-statistic 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 the supremum of a set of L-statistics with increasing weights. Hence, if one assigns larger weights to larger observations, a natural risk statistic becomes a law invariant coherent risk statistic. However, assigning larger weights to larger observations is not robust. Therefore, coherent risk measures are in general not robust.

7.2 Comparison with Insurance Risk Measures

Similarly, we can define the insurance risk statistics, the data-based versions of insurance risk measures, as follows:

Definition 3. A risk statistic $\hat{\rho} : \mathbb{R}^n \to \mathbb{R}$ is called an insurance risk statistic, if it satisfies the following Axiom 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 insurance risk statistics.

Theorem 5 *Let $x_{(1)}, ..., x_{(n)}$ be the order statistics of the data $\tilde{x}$ 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*

$$\hat{\rho}(\tilde{x}) = \sum_{i=1}^{n} w_i x_{(i)}, \ \forall \tilde{x} \in \mathbb{R}^n. \tag{17}$$

Proof. See Appendix D. *Q.E.D.*

Comparing (8) and (17), we see that a natural risk statistic is the supremum of a set 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 observation is specified. Hence, nature risk statistics incorporate the idea of evaluating risk under different scenarios, so do 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 $\tilde{y} = (3, 2, 4)$ and $\tilde{z} = (9, 4, 16)$. By simple calculation we have

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

even though $\tilde{y}$ and $\tilde{z}$ 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 $(0.72, 0.08, 0.2)$.

7.3 Comparison from the Viewpoint of Computation

We compare the computation of TCE and TCM in two aspects: whether it is easy to compute a risk measure from a regulator's viewpoint, and whether it is easy to incorporate a risk measure into portfolio optimization from an individual bank's viewpoint. First, the computation of TCM is at least as easy as the computation of TCE, since the former only involves the computation of quantile. Second, it is easier to do portfolio optimization with respect to TCE than to TCM, as the expectation 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 Rockafellar and Uryasev, 2000, 2002). 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 imposed by governmental regulations, should be a matter left for investors not the regulator.

8 Some Counterexamples Showing that Tail Conditional Median Satisfies Subadditivity

In the existing literature, some examples are used to show that 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 disappears.

Example 1. The VaRs in the example on page 216 of Artzner, Delbaen, Eber, and Heath (1999) are not correctly calculated. Actually in that example, the 1% VaR[15] 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 Artzner, Delbaen, Eber, and Heath (1999) shows that the 10% VaR does not satisfy subadditivity for X_1 and X_2. However, the 10% tail conditional 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 Dunkel and Weber (2007) shows that the 99% VaR of L_1 and L_2 are equal, although apparently L_2 is much more risky than L_1. However, tail conditional median at 99% level (or 99.5% VaR), of L_1 is equal to 10^{10}, which is much larger than 1, tail conditional median at 99% level (99.5% VaR) of L_2. In other words, if one looks at tail conditional median at 99% level, one can correctly compare the risk of the two portfolios.

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 prospect theory in psychology.

[15] In Artzner, Delbaen, Eber, and Heath (1999), 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 Artzner, Delbaen, Eber, and Heath (1999) corresponds to VaR at level $1-\alpha$ in this paper.

Comparing to existing risk measures, natural risk statistics include tail conditional median, which is more robust than tail conditional expectation suggested by coherent risk measures; and, unlike insurance risk measures, natural risk statistics can also incorporate scenario analysis. Natural risk statistics include VaR (with scenario analysis) as a special case and therefore show 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 regulation, and vice versa. For example, coherent and convex risk measures may be good for internal risk measurement, 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, 2001; Staum, 2004). However, as we point out, for external risk measures one may prefer a different set of properties, including consistency in implementation which means robustness.

A Proof of Theorem 1

The proof relies on the following two lemmas, which depend heavily on the properties of the interior points of the set $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \leq y_2 \leq \cdots \leq y_n\}$ and hence are only true for the interior points. 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 Axiom C1-C4 and $\hat{\rho}(\tilde{z}) = 1$ there exists a weight $\tilde{w} = (w_1, \ldots, w_n)$ such that the linear functional $\lambda(\tilde{x}) \triangleq \sum_{i=1}^n w_i x_i$ satisfies*

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

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

Proof. Let $U = \{\tilde{x} \mid \hat{\rho}(\tilde{x}) < 1\} \cap \mathcal{B}$. For any $\tilde{x}$, $\tilde{y} \in \mathcal{B}$, $\tilde{x}$ and $\tilde{y}$ are comonotonic. Then Axiom 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 $\lambda(\tilde{x}) \triangleq \sum_{i=1}^n w_i x_i$ 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 $\alpha_0 \in (0,1)$ 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 $\varepsilon > 0$, $\hat{\rho}(\tilde{x}^{\alpha_0} + \varepsilon\tilde{w}) < 1$, $\tilde{x}^{\alpha_0} + \varepsilon\tilde{w} \in U$. 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 $\lambda(\tilde{x}) < 1$ for all $\tilde{x}$ such that $\tilde{x} \in \mathcal{B}$ and $\hat{\rho}(\tilde{x}) < 1$, from which (19) holds. *Q.E.D.*

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 Axiom C1-C4, there exists a weight $\tilde{w} = (w_1, \ldots, w_n)$ such that*

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

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

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

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 $\lambda(\tilde{x}) \triangleq \sum_{i=1}^{n} w_i x_i$ 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 $\hat{\rho}(\tilde{x})\geq\lambda(\tilde{x})$, 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 $\lambda(\tilde{x})\triangleq\sum_{i=1}^n w_i x_i$, with $\tilde{w}=(w_1,\ldots,w_n)$ satisfying (26), (27) and $\hat{\rho}(\tilde{x})\geq\lambda(\tilde{x}),\forall\tilde{x}\in\mathcal{B}$, 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 $\hat{\rho}(\tilde{x})\geq\lambda(\tilde{x}),\forall\tilde{x}\in\mathcal{B}$, 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 $\lambda(\tilde{x})\triangleq\sum_{i=1}^n w_i x_i$ with $\tilde{w}=(w_1,\ldots,w_n)$ satisfying (26), (27) and $\hat{\rho}(\tilde{x})\geq\lambda(\tilde{x}),\forall\tilde{x}\in\mathcal{B}$, and $\hat{\rho}(\tilde{z}+b\mathbf{1})=\lambda(\tilde{z}+b\mathbf{1})$, or equivalently $\hat{\rho}(\tilde{x})\geq\lambda(\tilde{x}),\forall\tilde{x}\in\mathcal{B}$, and $\hat{\rho}(\tilde{z})=\lambda(\tilde{z})$. Thus, $\tilde{w}$ also satisfies (28). *Q.E.D.*

The proof for Theorem 1 is as follows.

Proof. (1) The proof of part (I). Suppose $\hat{\rho}$ is defined by (7), then obviously $\hat{\rho}$ satisfies Axiom 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,

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

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 $\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^n w_i^{(\tilde{y})} x_i$, $\forall \tilde{y} \in \mathcal{B}^o$, $\hat{\rho}(\tilde{x}) = \sum_{i=1}^n w_i^{(\tilde{x})} x_i$. 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 $\tilde{x}^0 = (x_1^0, \ldots, x_n^0)$ be any boundary point of $\mathcal{B}$. Then there exists a sequence $\{\tilde{x}^k = (x_1^k, \ldots, x_n^k)\}_{k=1}^\infty \subset \mathcal{B}^o$ 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

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

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

B Proof of Theorem 2

Proof. (I) (1) For $\forall \tilde{x} \leq 0$, Axiom C2 implies $\hat{\rho}(\tilde{x}) \leq \hat{\rho}(0) = 0$, 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 $0 \leq \tilde{x} - \alpha\mathbf{1}$. Axiom C2 and C1 imply that $\hat{\rho}(0) \leq \hat{\rho}(\tilde{x} - \alpha\mathbf{1}) = \hat{\rho}(\tilde{x}) - \alpha$. So $\hat{\rho}(\tilde{x}) \geq \alpha > 0$ and henceforth $\tilde{x} \notin \mathcal{A}_{\hat{\rho}}$, i.e., 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 $\hat{\rho}(\tilde{x}) \leq 0$, $\hat{\rho}(\tilde{y}) \leq 0$, and $\lambda\tilde{x}$ and $(1-\lambda)\tilde{y}$ are comonotonic for any $\lambda \in [0,1]$. Thus Axiom C3 implies

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

Hence $\lambda\tilde{x} + (1-\lambda)\tilde{y} \in \mathcal{A}_{\hat{\rho}}$, i.e., D3 holds. (4) For any $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$ and $a > 0$, we have $\hat{\rho}(\tilde{x}) \leq 0$ and Axiom C1 implies $\hat{\rho}(a\tilde{x}) = a\hat{\rho}(\tilde{x}) \leq 0$. Thus, $a\tilde{x} \in \mathcal{A}_{\hat{\rho}}$, i.e., D4 holds. (5) For any $\tilde{x} \leq \tilde{y}$ and $\tilde{y} \in \mathcal{A}_{\hat{\rho}}$, we have $\hat{\rho}(\tilde{y}) \leq 0$. By Axiom C2, $\hat{\rho}(\tilde{x}) \leq \hat{\rho}(\tilde{y}) \leq 0$. Hence $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$, i.e., D5 holds. (6) If $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$, then $\hat{\rho}(\tilde{x}) \leq 0$. For any permutation $(i_1, \ldots, i_n)$, Axiom C4 implies $\hat{\rho}((x_{i_1}, \ldots, x_{i_n})) = \hat{\rho}(\tilde{x}) \leq 0$. 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 $\hat{\rho}(\tilde{x}^k) \leq 0, \forall k$. Suppose the limit $\tilde{x} \notin \mathcal{A}_{\hat{\rho}}$. Then $\hat{\rho}(\tilde{x}) > 0$. There exists $\delta > 0$ such that $\hat{\rho}(\tilde{x} - \delta\mathbf{1}) > 0$. Since $\tilde{x}^k \to \tilde{x}$, it follows that there exists $K \in \mathbb{N}$ such that $\tilde{x}^K > \tilde{x} - \delta\mathbf{1}$. By Axiom C2, $\hat{\rho}(\tilde{x}^K) \geq \hat{\rho}(\tilde{x} - \delta\mathbf{1}) > 0$, which contradicts $\hat{\rho}(\tilde{x}^K) \leq 0$. So $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$, i.e., $\mathcal{A}_{\hat{\rho}}$ is closed.

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

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

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

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

where the second equality follows from Axiom 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 Axiom 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 Axiom 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 Axiom D3 that $\frac{1}{2}(\tilde{x} - m\mathbf{1}) + \frac{1}{2}(\tilde{y} - n\mathbf{1}) \in \mathcal{A}$. By Axiom D4, the previous formula implies $\tilde{x} + \tilde{y} - (m+n)\mathbf{1} \in \mathcal{A}$. Therefore, $\hat{\rho}_{\mathcal{A}}(\tilde{x} + \tilde{y}) \le m + n$. 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 $\hat{\rho}_{\mathcal{A}}(\tilde{x} + \tilde{y}) \le \hat{\rho}_{\mathcal{A}}(\tilde{x}) + \hat{\rho}_{\mathcal{A}}(\tilde{y})$. So C3 holds. (4) Fix any $\tilde{x} \in \mathbb{R}^n$ and any permutation $(i_1, \ldots, i_n)$. Then for any $m \in \mathbb{R}$, Axiom 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., C4 holds.

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

$$\begin{aligned}\hat{\rho}_{\mathcal{A}_{\hat{\rho}}}(\tilde{x}) &= \inf\{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{A}_{\hat{\rho}}\} = \inf\{m \mid \hat{\rho}(\tilde{x} - m\mathbf{1}) \le 0\} \\ &= \inf\{m \mid \hat{\rho}(\tilde{x}) \le m\} = \hat{\rho}(\tilde{x}),\end{aligned}$$

where the third equality follows from Axiom C1.

(IV) For any $\tilde{x} \in \mathcal{D}$, we have $\hat{\rho}_{\mathcal{D}}(\tilde{x}) \le 0$. 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 $\hat{\rho}_{\mathcal{D}}(\tilde{x}) \le 0$, i.e., $\inf\{m \mid \tilde{x} - m\mathbf{1} \in \mathcal{D}\} \le 0$. 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 $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}} = \bar{\mathcal{D}}$. *Q.E.D.*

C Proof of Theorem 4

The proof for theorem 4 follows the same line as the proof for Theorem 1. We first prove two lemmas that are similar to Lemma 1 and 2.

Lemma 3 *Let $\mathcal{B} = \{\tilde{y} \in \mathbb{R}^n \mid y_1 \le y_2 \le \cdots \le y_n\}$. For any fixed $\tilde{z} \in \mathcal{B}$ and any $\hat{\rho}$ satisfying Axiom C1, C2, E3, C4 and $\hat{\rho}(\tilde{z}) = 1$, there exists a weight $\tilde{w} = (w_1, \ldots, w_n)$ satisfying* (14) *such that the linear functional $\lambda(\tilde{x}) \triangleq \sum_{i=1}^n w_i x_i$ satisfies*

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

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

Proof. Let $U = \{\tilde{x} \mid \hat{\rho}(\tilde{x}) < 1\}$. Axiom 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:

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

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

then $\tilde{w}$ satisfies condition (14).

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, Axiom 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 $\lambda(\tilde{x}) < 1$ for all $\tilde{x}$ such that $\hat{\rho}(\tilde{x}) < 1$, from which (36) holds. *Q.E.D.*

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 Axiom C1, C2, E3 and C4, there exists a weight $\tilde{w} = (w_1, \ldots, w_n)$ satisfying* (12), (13) *and* (14), *such that*

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

Proof. The proof is obtained by using Lemma 3 and following the same line as the proof for Lemma 2. *Q.E.D.*

The proof for Theorem 4 is as follows.

Proof. (1) Proof for part (I). We only need to show that under condition (14), the risk statistic (15) 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 $\mathcal{W} = \{\tilde{w} = (w_1,\ldots,w_n)\} \subset \mathbb{R}^n$ with each $\tilde{w} \in \mathcal{W}$ satisfying (12), (13) and (14), 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 (12), (13) and (14) 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 $\hat{\rho}(\tilde{x}) \ge \sum_{i=1}^n w_i^{(\tilde{y})} x_i$, $\forall \tilde{y} \in \mathcal{B}$; 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. *Q.E.D.*

D Proof of Theorem 5

Proof. (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 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 third equality follows from the comonotonic additivity. Therefore, by (48)

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

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. *Q.E.D.*

E Calculate Influence Functions of TCE and TCM

Proposition 1 *Suppose the random loss X has a density $f_X(\cdot)$ that is continuous and positive at $VaR_{\frac{1+\alpha}{2}}(X)$, then the influence function of TCM_α is given by*

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

Suppose the random loss X has a density $f_X(\cdot)$ that is continuous and positive at $VaR_\alpha(X)$, then the influence function of TCE_α is given by

$$IF(x, TCE_\alpha, X) = \begin{cases} VaR_\alpha(X) - TCE_\alpha(X), & \text{if } x \leq VaR_\alpha(X), \\ \frac{x}{1-\alpha} - TCE_\alpha(X) - \frac{\alpha}{1-\alpha} VaR_\alpha(X), & \text{if } x > VaR_\alpha(X). \end{cases} \tag{56}$$

Proof. The result for TCM is given by equation (3.2.3) in Staudte and Sheather (1990). To show (56), note that by equation (3.2.4) in Staudte and Sheather (1990) the influence function of the $(1-\alpha)$-trimmed mean $T_{1-\alpha}(X) \triangleq E[X|X < \mathrm{VaR}_\alpha(X)]$ is

$$IF(x, T_{1-\alpha}, X) = \begin{cases} \frac{x-(1-\alpha)\mathrm{VaR}_\alpha(X)}{\alpha} - E[X|X < \mathrm{VaR}_\alpha(X)], & \text{if } x \le \mathrm{VaR}_\alpha(X), \\ \mathrm{VaR}_\alpha(X) - E[X|X < \mathrm{VaR}_\alpha(X)], & \text{if } x > \mathrm{VaR}_\alpha(X). \end{cases} \tag{57}$$

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

$$IF(x, E[X], X) = x - E[X]. \tag{58}$$

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

$$IF(x, E[X], X) = \alpha IF(x, T_{1-\alpha}, X) + (1-\alpha) IF(x, \mathrm{TCE}_\alpha, X). \tag{59}$$

Now (56) follows from equations (57), (58) and (59). *Q.E.D.*

References

Acerbi, C., Nordio, C., Sirtori, C., 2001. Expected shortfall as a tool for financial risk management. Working paper. Abaxbank, Italy.

Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* **26(7)**, 1487–1503.

Ahmed, S., Filipović, D., and Svindland, G., 2008. A note on natural risk statistics. *Operations Research Letters* **36**, 662–664.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. *Mathematical Finance* **9(3)**, 203–228.

Basel Committee on Banking Supervision, 2006. *International Convergence of Capital Measurement and Capital Standards: A Revised Framework (Comprehensive Version).*

Brumelle, S. L., 1974. When does diversification between two investments pay? *Journal of Financial and Quantitative Analysis* **9(3)**, 473–482.

Daníelsson, J., Jorgensen, B. N., Samorodnitsky, G., Sarma, M., de Vries, C. G., 2005. Subadditivity re-examined: the case for Value-at-Risk. Working paper. London School of Economics. http://ideas.repec.org/p/fmg/fmgdps/dp549.html.

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

Denneberg, D., 1994. *Non-Additive Measure and Integral*. Kluwer Academic Publishers, Boston.

Dhaene, J., Goovaerts, M. J., Kaas, R., 2003. Economic capital allocation derived from risk measures. *North American Actuarial Journal* **7(2)**, 44–59.

Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., Tang, Q., Vyncke, D., 2006. Risk measures and comonotonicity: a review. *Stochastic Models* **22(4)**, 573–606.

Duffie, D., Epstein, L. G., 1992. Stochastic differential utility. *Econometrica*, **60(2)**, 353–394.

Dunkel, J., Weber, S., 2007. Efficient Monte Carlo methods for convex risk measures in portfolio credit risk models. Working paper. Max-Planck-Institute.

Eeckhoudt, L., Schlesinger, H., 2006. Putting risk in its proper place. *American Economic Review* **96(1)**, 280–289.

Eling, M., Tibiletti, L., 2008. Internal vs. external risk measures: how capital requirements differ in practice. Working paper. University of St. Gallen.

Epstein, L. G., Schneider, M., 2003. Recursive multiple-priors. *Journal of Economic Theory* **113(1)**, 1–31.

Epstein, L. G., Zin, S. E., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica* **57(4)**, 937–969.

Föllmer, H., Schied, A., 2002. Convex measures of risk and trading constraints. *Finance and Stochastics* **6(4)**, 429-447.

Frittelli, M., Gianin, E. R., 2002. Putting order in risk measures. *Journal of Banking and Finance* **26(7)**, 1473–1486.

Garcia, R., Renault, É., Tsafack, G., 2007. Proper conditioning for coherent VaR in portfolio management. *Management Science* **53(3)**, 483–494.

Hampel, F. R., 1974. The influence curve and its role in robust estimation. *Journal of American Statistical Association* **69**, 383–393.

Hart, H. L. A., 1994. *The Concept of Law*, 2nd ed. Clarendon Press, Oxford.

Heyde, C. C., Kou, S. G., 2004. On the controversy over tailweight of distributions. *Operations Research Letters* **32(5)**, 399–408.

Hong, C.-S., Herk, L. F., 1996. Increasing risk aversion and diversification preference. *Journal of Economic Theory* **70(1)**, 180–200.

Huber, P. J., 1981. *Robust Statistics*. Wiley, New York.

Ibragimov, R., Walden, J., 2006. The limits of diversification when losses may be large. Harvard Institute of Economic Research Discussion Paper No. 2104. http://ssrn.com/abstract=880081.

Institute of Transportation Engineers, 2000. Speeding zoning information. http://-www.ite.org/standards/speed_zoning.pdf.

Jaschke, S., Küchler, U., 2001. Coherent risk measures and good deal bounds. *Finance and Stochastics* **5(2)**, 181–200.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* **47(2)**, 263–291.

Kijima, M., 1997. The generalized harmonic mean and a portfolio problem with dependent assets. *Theory and Decision* **43(1)**, 71–87.

Koopmans, T. C., 1960. Stationary ordinal utility and impatience. *Econometrica* **28(2)**, 287–309.

Kusuoka, S., 2001. On law invariant coherent risk measures. *Advances in Mathematical Economics* **3**, 83–95.

McMinn, R. D., 1984. A general diversification theorem: a note. *Journal of Finance* **39(2)**, 541–550.

Pflug, G., 2000. Some remarks on the Value-at-Risk and the Conditional Value-at-Risk. In: Uryasev, S.(Ed.), *Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer Academic Publishers, 278–287.

Quiggin, J. C., 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization* **3(4)**, 323–343.

Rockafellar, R. T., Uryasev, S., 2000. Optimization of Conditional Value-At-Risk. *Journal of Risk* **2(3)**, 21–41.

Rockafellar, R. T., Uryasev, S., 2002. Conditional Value-at-Risk for general loss distributions. *Journal of Banking and Finance* **26(7)**, 1443–1471.

Samuelson, P. A., 1967. General proof that diversification pays. *The Journal of Financial and Quantitative Analysis* **2(1)**, 1–13.

Schmeidler, D., 1986. Integral representation without additivity. *Proceedings of the American Mathematical Society* **97(2)**, 255–261.

Schmeidler, D., 1989. Subjective probability and expected utility without additivity. *Econometrica* **57(3)**, 571–587.

Song, Y., Yan, J.-A., 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.

Song, Y., Yan, J.-A., 2006. Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. Working paper. Chinese Academy of Sciences.

Staudte, R. G., Sheather, S. J., 1990. *Robust Estimation and Testing*. Wiley, New York.

Staum, J., 2004. Fundamental theorems of asset pricing for good deal bounds. *Mathematical Finance* **14(2)**, 141–161.

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

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5(4)**, 297–323.

Wakker, P., Tversky, A., 1993. An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty* **7(2)**, 147–175.

Wang, S. S., Young, V. R., Panjer, H. H., 1997. Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics* **21(2)**, 173–183.

Wang, T., 2003. Conditional preferences and updating. *Journal of Economic Theory*, **108(2)**, 286–321.

Yaari, M. E., 1987. The dual theory of choice under risk. *Econometrica* **55(1)**, 95–115.