

# External Risk Measures and Basel Accords

Steven Kou

Department of IEOR, 312 Mudd Building, Columbia University, New York,  
New York 10027, sk75@columbia.edu

Xianhua Peng

Department of Mathematics, The Hong Kong University of Science and Technology,  
Kowloon, Hong Kong, maxhpeng@ust.hk

Chris C. Heyde

Deceased

Choosing a proper external risk measure is of great regulatory importance, as exemplified in the Basel II and Basel III Accord which use Value-at-Risk (VaR) with scenario analysis as the risk measures for setting capital requirements. We argue a good external risk measure should be robust with respect to model misspecification and small changes in the data. A new class of data-based risk measures called natural risk statistics are proposed to incorporate robustness. Natural risk statistics are characterized by a new set of axioms; they include the Basel II and III risk measures and a subclass of robust risk measures as special cases; therefore, they provide a theoretical framework for understanding and, if necessary, extending the Basel accords.

*Key words:* financial regulation; capital requirements; risk measure; scenario analysis; robustness

*MSC2000 Subject Classification:* Primary: 91B30, 62P20; Secondary: 91B08

*OR/MS subject classification:* Primary: Risk, regulations

---

**1. Introduction.** Broadly speaking, a risk measure attempts to assign a single numerical value to the random loss of a portfolio of assets. Mathematically, let  $\Omega$  be the set of all the possible states of nature at the end of an observation period, and  $\mathcal{X}$  be the set of financial losses, which are random variables defined on  $\Omega$ . Then a risk measure  $\rho$  is a mapping from  $\mathcal{X}$  to the real line  $\mathbb{R}$ . Obviously, it can be problematic to use one number to summarize the whole statistical distribution of the potential loss. Therefore, one should avoid doing this if it is at all possible. In many cases, however, there is no other choice. Examples of such cases include margin requirements in financial trading, insurance premiums, and regulatory capital requirements. Consequently, choosing a good risk measure becomes a problem of great practical importance.

The Basel accord risk measures are used for setting capital requirements for the banking books and trading books of financial institutions. Since the Basel accord risk measures lead to important regulations, there are a lot of debates on what risk measures are good in the finance industry. In fact, one can even question whether it is efficient to set up capital requirements using any risk measures. For example, in an interesting paper Keppo, Kofman, and Meng [31] analyze the effect of the Basel accord capital requirements on the behavior of a bank and show surprisingly that imposing trading book capital requirements may in fact postpone recapitalization of the bank and hence increase its default probability.

One of the most widely used risk measures is *Value-at-Risk* (*VaR*), which is a quantile at some pre-defined probability level. More precisely, let  $F(\cdot)$  be the distribution function of the random loss  $X$ , then for a given  $\alpha \in (0, 1)$ , VaR of  $X$  at level  $\alpha$  is defined as  $\text{VaR}_\alpha(X) := \inf\{x \mid F(x) \geq \alpha\} = F^{-1}(\alpha)$ . In practice,  $\text{VaR}_\alpha(X)$  is usually estimated from a sample of  $X$ , i.e., a data set  $\tilde{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ .

Gordy [18] provides a theoretical foundation for the Basel accord banking book risk measure by demonstrating that under certain conditions the risk measure is asymptotically equivalent to the 99.9% VaR. The Basel II and Basel III risk measures for trading books [5, 7] are both special cases of *VaR with scenario analysis*, which is a class of risk measures involving calculation and comparison of VaR under different scenarios; each scenario refers to a specific economic regime such as an economic boom and a financial crisis. The loss distributions under different scenarios are substantially different and hence the values of VaR calculated under different scenarios are distinct from each other; for example, the VaR calculated under the scenario of the 2008 financial crisis is much higher than the VaR calculated under a scenario corresponding to normal market conditions. The exact formulae of the Basel II and Basel III risk measures are given in Section 4.

Although the Basel II and Basel III risk measures for trading books are of great regulatory importance, there has been *no* axiomatic justification for their use. 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 the actuarial literature such as Wang, Young, and Panjer [48], we shall define a new class of risk measures that satisfy a new set of axioms. The new class of risk measures include VaR with scenario analysis, and particularly the Basel II and Basel III risk measures, as special cases. Thus, we provide a theoretical framework for understanding and extending the Basel accords when needed. Indeed, the framework includes as special cases some proposals to address the procyclicality problem in Basel II such as the counter-cyclical indexing risk measure suggested by Gordy and Howells [19].

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 internal risk management at individual institutions, and *external* risk measures used for external regulation and imposed for all the relevant institutions. The differences between internal and external risk measures mirror the differences between internal standards (such as morality) and external standards (such as law and regulation). Internal risk measures are applied in the interest of an institution’s shareholders or managers, while external risk measures are used by regulatory agencies to maintain safety and soundness of the financial system. A risk measure may be suitable for internal management but not for external regulation, or vice versa.

In this paper, we shall focus on external risk measures from the viewpoint of regulatory agencies. In particular, we emphasize that an external risk measure should be robust (see Section 5).

The main results of the paper are as follows: (i) We postulate a new set of axioms and define a new class of risk measures called natural risk statistics; furthermore, we give two complete characterizations of natural risk statistics (Section 3.2). (ii) We show that natural risk statistics include the Basel II and Basel III risk measures as special cases and thus provide an axiomatic framework for understanding and, if necessary, extending them (Section 4). (iii) We completely characterize data-based coherent risk measures and show that *no* coherent risk measure is robust with respect to small changes in the data (Section 3.3 and 5.6). (iv) We completely characterize data-based insurance risk measures and show that *no* insurance risk measure is robust with respect to model misspecification (Section 3.4 and 5.6). (v) We argue that an external risk measure should be robust, motivated by philosophy of law and issues in external regulations (Section 5). (vi) We show that tail conditional median, a special case of natural risk statistics, is more robust than tail conditional expectation suggested by coherent risk measures (Section 5.4). (vii) We show that natural risk statistics include a subclass of robust risk measures that are suitable for external regulation (Section 5.5). (viii) We provide other critiques of the subadditivity axiom of coherent risk measures from the viewpoints of diversification and bankruptcy protection (Section 6). (ix) We derive the Euler capital allocation rule under a subclass of natural risk statistics including the Basel II and III risk measure (Section 7).

## 2. Review of Existing Risk Measures.

**2.1 Coherent and Convex Risk Measures.** Artzner, Delbaen, Eber, and Heath [4] propose the coherent risk measures that satisfy the following three axioms:

**Axiom A1.** Translation invariance and positive homogeneity:  $\rho(aX + b) = a\rho(X) + b$ ,  $\forall a \geq 0, \forall b \in \mathbb{R}, \forall X \in \mathcal{X}$ .

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

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

Axiom A1 states that the risk of a financial position is proportional to its size, and a sure loss of amount  $b$  simply increases the risk by  $b$ . Axiom A1 is proposed from the accounting viewpoint. For external risk measures such as those used for setting margin deposits and capital requirements, the accounting-based axiom seems to be reasonable. Axiom A2 is a minimum requirement for a reasonable risk measure. What is questionable lies in Axiom A3, which basically means that “a merger does not create extra risk” (see Artzner et al. [4, p. 209]). We will discuss the controversies related to this axiom in Section 6. Artzner et al. [4] and Delbaen [9] also present an equivalent approach for defining coherent risk measures via acceptance sets. Föllmer and Schied [13] and Frittelli and Gianin [14] propose the convex risk measures that relax Axiom A1 and Axiom A3 to a single convexity axiom:  $\rho(\lambda X + (1 - \lambda)Y) \leq$

$$\lambda\rho(X) + (1 - \lambda)\rho(Y), \quad \forall X, Y \in \mathcal{X}, \forall \lambda \in [0, 1].$$

A risk measure  $\rho$  is coherent if and only if there exists a family  $\mathcal{Q}$  of probability measures such that  $\rho(X) = \sup_{Q \in \mathcal{Q}} \{E^Q[X]\}$ ,  $\forall X \in \mathcal{X}$ , where  $E^Q[X]$  is the expectation of  $X$  under the probability measure  $Q$  (see Huber [24], Artzner et al. [4], and Delbane [9]). Each  $Q \in \mathcal{Q}$  can be viewed as a prior probability, so measuring risk by a coherent risk measure amounts to computing the maximal expectation under a set of prior probabilities. Coherent and convex risk measures are closely connected to the good deal bounds of asset prices in incomplete markets (see, e.g., Jäschke and Kückler [29], Staum [45]).

Artzner et al. [4] suggest using a specific coherent risk measure called tail conditional expectation (TCE). For a random loss  $X$  with a continuous distribution, TCE of  $X$  at level  $\alpha$  is defined as

$$\text{TCE}_\alpha(X) := E[X|X \geq \text{VaR}_\alpha(X)]. \quad (1)$$

For  $X$  with a general probability distribution,  $\text{TCE}_\alpha(X)$  is defined as a regularized version of  $E[X|X \geq \text{VaR}_\alpha(X)]$  (Rockafellar and Uryasev [38]). Tail conditional expectation is also called expected shortfall (Acerbi and Tasche [1]) or conditional value-at-risk (Rockafellar and Uryasev [38]).

A risk measure is called a law-invariant coherent risk measure (Kusuoka [33]) if it satisfies Axiom A1-A3 and the following Axiom A4:

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

Insisting on a coherent or convex risk measure rules out the use of VaR, for VaR does not universally satisfy subadditivity or convexity. The exclusion of VaR gives rise to a serious inconsistency between academic theories and governmental practices. By requiring subadditivity only for comonotonic random variables, we will define a new class of risk measures that include VaR and, more importantly, VaR with scenario analysis, thus eliminating the inconsistency (see Section 3).

**2.2 Insurance Risk Measures.** Wang, Young, and Panjer [48] propose the insurance risk measures that satisfy the following 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  $\omega_1$  and  $\omega_2$  in  $\Omega$ .)

**Axiom B4.** Continuity:  $\lim_{d \rightarrow 0} \rho((X - d)^+) = \rho(X^+)$ ,  $\lim_{d \rightarrow -\infty} \rho(\max(X, d)) = \rho(X)$ , and  $\lim_{d \rightarrow \infty} \rho(\min(X, d)) = \rho(X)$ ,  $\forall X$ , where  $x^+ := \max(x, 0)$ ,  $\forall x \in \mathbb{R}$ .

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

Comonotonic random variables are studied by Yaari [49], Schmeidler [40], Denneberg [10], and others. If two random variables  $X$  and  $Y$  are comonotonic,  $X(\omega)$  and  $Y(\omega)$  always move in the same direction however the state  $\omega$  changes. For example, the payoffs of a call option and its underlying asset are comonotonic.

Wang et al. [48] show that  $\rho$  is an insurance risk measure if and only if  $\rho$  has a Choquet integral representation with respect to a distorted probability:

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

where  $g(\cdot)$  is called the distortion function which is nondecreasing and satisfies  $g(0) = 0$  and  $g(1) = 1$ .  $g \circ P$  is called the distorted probability and defined by  $g \circ P(A) := g(P(A))$  for any event  $A$ . In general, an insurance risk measure does not satisfy subadditivity unless  $g(\cdot)$  is concave (Denneberg [10]). 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 priors.

VaR with scenario analysis, such as the Basel II and Basel III risk measures (see Section 4 for their definition), are not insurance risk measures, although VaR itself is an insurance risk measure. The main reason that insurance risk measures cannot incorporate scenario analysis or multiple priors is that they

require comonotonic additivity. Wang et al. [48] impose comonotonic additivity based on the argument that comonotonic random variables do not hedge against each other. However, comonotonic additivity holds only if a single prior is considered. If multiple priors are considered, one can get strict subadditivity rather than additivity for comonotonic random variables. Hence, Axiom B3 may be too restrictive. To incorporate multiple priors, we shall relax the comonotonic additivity to comonotonic subadditivity (see Section 3).

The mathematical concept of comonotonic subadditivity is also studied independently by Song and Yan [42], who give a representation of the functionals satisfying comonotonic subadditivity or comonotonic convexity from a mathematical perspective. Song and Yan [43] give a representation of risk measures that respect stochastic orders and are comonotonically subadditive or convex. There are several major differences between their work and this paper: (i) The new risk measures proposed in this paper are different from those considered in Song and Yan [42, 43]. In particular, the new risk measures include VaR with scenario analysis, such as the Basel II and Basel III risk measures, as special cases. However, VaR with scenario analysis are not included in the class of risk measures considered by Song and Yan [42, 43]. (ii) The framework of Song and Yan [42, 43] is based on subjective probability models, but the framework of the new risk measures is explicitly based on data and scenario analysis (Section 3.1). (iii) We provide legal and economic reasons for postulating the comonotonic subadditivity axiom (Section 5 and 6). (iv) We provide two complete characterizations of the new risk measures (Section 3.2). (v) We completely characterize the data-based coherent and insurance risk measures so that we can compare them with the new risk measures (Section 3.3 and 3.4).

### 3. Natural Risk Statistics.

**3.1 Risk Statistics: Data-based Risk Measures.** In external regulation, the behavior of the random loss  $X$  under different scenarios is preferably represented by different sets of data observed or generated under those scenarios because specifying accurate models for  $X$  (under different scenarios) is usually very difficult. More precisely, suppose the behavior of  $X$  is represented by a collection of data  $\tilde{x} = (\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^m) \in \mathbb{R}^n$ , where  $\tilde{x}^i = (x_1^i, \dots, x_{n_i}^i) \in \mathbb{R}^{n_i}$  is the data subset that corresponds to the  $i$ -th scenario and  $n_i$  is the sample size of  $\tilde{x}^i$ ;  $n_1 + n_2 + \dots + n_m = n$ . For each  $i = 1, \dots, m$ ,  $\tilde{x}^i$  can be a data set based on historical observations, hypothetical samples simulated according to a model, or a mixture of observations and simulated samples.  $X$  can be either discrete or continuous. For example, the data used in the calculation of the Basel III risk measure comprise 120 data subsets corresponding to 120 different scenarios ( $m = 120$ ); see Section 4 for the details of the Basel III risk measures.

A *risk statistic*  $\hat{\rho}$  is simply a mapping from  $\mathbb{R}^n$  to  $\mathbb{R}$ . It is a data-based risk measure that maps  $\tilde{x}$ , the data representation of the random loss  $X$ , to  $\hat{\rho}(\tilde{x})$ , the risk measurement of  $X$ . In this paper, we will define a new set of axioms for risk statistics instead of risk measures because (i) risk statistics can directly measure risk from observations without specifying subjective models, which greatly reduces model misspecification error; (ii) risk statistics can incorporate forward-looking views or prior knowledge by including data subsets generated by models based on such views or knowledge; and (iii) risk statistics can incorporate multiple prior probabilities on the set of scenarios which reflect multiple beliefs about the probabilities of occurrence of different scenarios.

**3.2 Axioms and a Representation of Natural Risk Statistics.** First, we define the notion of scenario-wise comonotonicity for two sets of data, which is the counterpart of the notion of comonotonicity for two random variables.  $\tilde{x} = (\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^m) \in \mathbb{R}^n$  and  $\tilde{y} = (\tilde{y}^1, \tilde{y}^2, \dots, \tilde{y}^m) \in \mathbb{R}^n$  are *scenario-wise comonotonic* if for  $\forall i, \forall 1 \leq j, k \leq n_i$ , it holds that  $(x_j^i - x_k^i)(y_j^i - y_k^i) \geq 0$ . Let  $\tilde{x}$  and  $\tilde{y}$  represent the observations of random losses  $X$  and  $Y$  respectively, then  $\tilde{x}$  and  $\tilde{y}$  are scenario-wise comonotonic means that  $X$  and  $Y$  move in the same direction under each scenario  $i$ ,  $i = 1, \dots, m$ , which is consistent with the notion that  $X$  and  $Y$  are comonotonic.

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

**Axiom C1.** Positive homogeneity and translation scaling:  $\hat{\rho}(a\tilde{x} + b\mathbf{1}) = a\hat{\rho}(\tilde{x}) + sb$ ,  $\forall \tilde{x} \in \mathbb{R}^n, \forall a \geq 0, \forall b \in \mathbb{R}$ , where  $s > 0$  is a scaling constant,  $\mathbf{1} = (1, 1, \dots, 1) \in \mathbb{R}^n$ .

**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  $x_j^i \leq y_j^i, j = 1, \dots, n_i; i = 1, \dots, m$ .

These two axioms (with  $s = 1$  in Axiom C1) are the counterparts of Axiom A1 and A2 for coherent

risk measures. Axiom C1 clearly yields  $\hat{\rho}(0 \cdot \mathbf{1}) = 0$  and  $\hat{\rho}(b\mathbf{1}) = sb$ , for any  $b \in \mathbb{R}$ , and Axioms C1 and C2 imply that  $\hat{\rho}$  is continuous. Indeed, suppose  $\hat{\rho}$  satisfies Axiom C1 and C2. Then for any  $\tilde{x} \in \mathbb{R}^n$ ,  $\varepsilon > 0$ , and  $\tilde{y} \in \mathbb{R}^n$  satisfying  $\tilde{x} - \varepsilon\mathbf{1} < \tilde{y} < \tilde{x} + \varepsilon\mathbf{1}$ , by 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  $\hat{\rho}(\tilde{x}) - s\varepsilon \leq \hat{\rho}(\tilde{y}) \leq \hat{\rho}(\tilde{x}) + s\varepsilon$ , which establishes the continuity of  $\hat{\rho}$ .

**Axiom C3.** Scenario-wise comonotonic subadditivity:  $\hat{\rho}(\tilde{x} + \tilde{y}) \leq \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y})$ , for any  $\tilde{x}$  and  $\tilde{y}$  that are scenario-wise comonotonic.

Axiom C3 relaxes the subadditivity requirement, Axiom A3, in coherent risk measures so that subadditivity is only required for comonotonic random variables. It also relaxes the comonotonic additivity requirement, Axiom B3, in insurance risk measures. In other words, if one believes either Axiom A3 or Axiom B3, then one has to believe the new Axiom C3.

**Axiom C4.** Empirical law invariance:

$$\hat{\rho}((\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^m)) = \hat{\rho}((x_{p_{1,1}}^1, \dots, x_{p_{1,n_1}}^1, x_{p_{2,1}}^2, \dots, x_{p_{2,n_2}}^2, \dots, x_{p_{m,1}}^m, \dots, x_{p_{m,n_m}}^m))$$

for any permutation  $(p_{i,1}, \dots, p_{i,n_i})$  of  $(1, 2, \dots, n_i)$ ,  $i = 1, \dots, m$ .

This axiom is the counterpart of the law invariance Axiom A4. It means that if two data  $\tilde{x}$  and  $\tilde{y}$  have the same empirical distributions under each scenario, i.e., the same order statistics under each scenario, then  $\tilde{x}$  and  $\tilde{y}$  should give the same measurement of risk.

A risk statistic  $\hat{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}$  is called a *natural risk statistic* if it satisfies Axiom C1-C4. The following theorem completely characterizes natural risk statistics.

**THEOREM 3.1** (i) For a given constant  $s > 0$  and an arbitrarily given set of weights  $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^n$  with each  $\tilde{w} = (w_1^1, \dots, w_{n_1}^1, \dots, w_1^m, \dots, w_{n_m}^m) \in \mathcal{W}$  satisfying the following conditions

$$\sum_{j=1}^{n_1} w_j^1 + \sum_{j=1}^{n_2} w_j^2 + \dots + \sum_{j=1}^{n_m} w_j^m = 1, \quad (3)$$

$$w_j^i \geq 0, j = 1, \dots, n_i; i = 1, \dots, m, \quad (4)$$

define a risk statistic  $\hat{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}$  as follows:

$$\hat{\rho}(\tilde{x}) := s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{j=1}^{n_1} w_j^1 x_{(j)}^1 + \sum_{j=1}^{n_2} w_j^2 x_{(j)}^2 + \dots + \sum_{j=1}^{n_m} w_j^m x_{(j)}^m \right\}, \quad \forall \tilde{x} = (\tilde{x}^1, \dots, \tilde{x}^m) \in \mathbb{R}^n, \quad (5)$$

where  $(x_{(1)}^i, \dots, x_{(n_i)}^i)$  is the order statistics of  $\tilde{x}^i = (x_1^i, \dots, x_{n_i}^i)$  with  $x_{(n_i)}^i$  being the largest,  $i = 1, \dots, m$ . Then the  $\hat{\rho}$  defined in (5) is a natural risk statistic.

(ii) If  $\hat{\rho}$  is a natural risk statistic, then there exists a set of weights  $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^n$  such that each  $\tilde{w} = (w_1^1, \dots, w_{n_1}^1, \dots, w_1^m, \dots, w_{n_m}^m) \in \mathcal{W}$  satisfies condition (3) and (4), and

$$\hat{\rho}(\tilde{x}) = s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{j=1}^{n_1} w_j^1 x_{(j)}^1 + \sum_{j=1}^{n_2} w_j^2 x_{(j)}^2 + \dots + \sum_{j=1}^{n_m} w_j^m x_{(j)}^m \right\}, \quad \forall \tilde{x} = (\tilde{x}^1, \dots, \tilde{x}^m) \in \mathbb{R}^n. \quad (6)$$

**PROOF.** See Appendix A. □

The main difficulty in proving Theorem 3.1 lies in part (ii). Axiom C3 implies that  $\hat{\rho}$  satisfies subadditivity on scenario-wise comonotonic sets of  $\mathbb{R}^n$ , such as the set  $\mathcal{B} := \{\tilde{y} = (\tilde{y}^1, \dots, \tilde{y}^m) \in \mathbb{R}^n \mid y_1^1 \leq y_2^1 \leq \dots \leq y_{n_1}^1; \dots; y_1^m \leq y_2^m \leq \dots \leq y_{n_m}^m\}$ . However, unlike the case of coherent risk measures, the existence of a set of weights  $\mathcal{W}$  that satisfies (6) does not follow easily from the proof developed by Huber [24]. The main difference here is that the 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 do and treating them involves greater effort. In particular, one should be very cautious when using the results of separating hyperplanes. For the case of  $m = 1$  (one scenario), Ahmed, Filipović, and Svindland [3] provide alternative shorter proofs for Theorem 3.1 and Theorem 3.3 using convex duality theory after seeing the first version of this paper.

Natural risk statistics can also be characterized via acceptance sets, as in the case of coherent risk measures. We show in Appendix B that for a natural risk statistic  $\hat{\rho}$ , the risk measurement  $\hat{\rho}(\tilde{x})$  is equal

to the minimum amount of cash that has to be added to the position corresponding to  $\tilde{x}$  to make the modified position acceptable.

**3.3 Comparison with Coherent Risk Measures.** To formally compare natural risk statistics with coherent risk measures, we first define the coherent risk statistics, the data-based versions of coherent risk measures. A risk statistic  $\hat{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}$  is called a *coherent risk statistic* if it satisfies Axiom C1, C2, and the following Axiom E3.

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

**THEOREM 3.2** *A risk statistic  $\hat{\rho}$  is a coherent risk statistic if and only if there exists a set of weights  $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^n$  such that each  $\tilde{w} \in \mathcal{W}$  satisfies (3) and (4), and*

$$\hat{\rho}(\tilde{x}) = s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{j=1}^{n_1} w_j^1 x_j^1 + \sum_{j=1}^{n_2} w_j^2 x_j^2 + \cdots + \sum_{j=1}^{n_m} w_j^m x_j^m \right\}, \quad \forall \tilde{x} = (\tilde{x}^1, \dots, \tilde{x}^m) \in \mathbb{R}^n. \quad (7)$$

**PROOF.** The proof for the “if” part is trivial. To prove the “only if” part, suppose  $\hat{\rho}$  is a coherent risk statistic. Let  $\Theta = \{\theta_1, \dots, \theta_n\}$  be a set with  $n$  elements and  $\mathcal{Z}$  be the set of all real-valued functions defined on  $\Theta$ . Define the functional  $E^*(Z) := \frac{1}{s} \hat{\rho}(Z(\theta_1), Z(\theta_2), \dots, Z(\theta_n))$ ,  $\forall Z \in \mathcal{Z}$ . By Axiom C1, C2, and E3,  $E^*(\cdot)$  satisfies the conditions in Proposition 10.1 on p. 252 of Huber and Ronchetti [25], so the result follows by applying that proposition.  $\square$

Natural risk statistics satisfy empirical law invariance, which coherent risk statistics do not. To better compare natural risk statistics and coherent risk measures, we define empirical-law-invariant coherent risk statistics, which are the counterparts of law-invariant coherent risk measures. A risk statistic  $\hat{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}$  is called an *empirical-law-invariant* coherent risk statistic if it satisfies Axiom C1, C2, E3, and C4. The following theorem completely characterizes empirical-law-invariant coherent risk statistics.

**THEOREM 3.3** (i) *For a given constant  $s > 0$  and an arbitrarily given set of weights  $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^n$  with each  $\tilde{w} = (w_1^1, \dots, w_{n_1}^1, \dots, w_1^m, \dots, w_{n_m}^m) \in \mathcal{W}$  satisfying the following conditions*

$$\sum_{j=1}^{n_1} w_j^1 + \sum_{j=1}^{n_2} w_j^2 + \cdots + \sum_{j=1}^{n_m} w_j^m = 1, \quad (8)$$

$$w_j^i \geq 0, j = 1, \dots, n_i; i = 1, \dots, m, \quad (9)$$

$$w_1^i \leq w_2^i \leq \cdots \leq w_{n_i}^i, i = 1, \dots, m, \quad (10)$$

define a risk statistic

$$\hat{\rho}(\tilde{x}) := s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{j=1}^{n_1} w_j^1 x_{(j)}^1 + \sum_{j=1}^{n_2} w_j^2 x_{(j)}^2 + \cdots + \sum_{j=1}^{n_m} w_j^m x_{(j)}^m \right\}, \quad \forall \tilde{x} = (\tilde{x}^1, \dots, \tilde{x}^m) \in \mathbb{R}^n, \quad (11)$$

where  $(x_{(1)}^i, \dots, x_{(n_i)}^i)$  is the order statistics of  $\tilde{x}^i = (x_1^i, \dots, x_{n_i}^i)$  with  $x_{(n_i)}^i$  being the largest,  $i = 1, \dots, m$ . Then the  $\hat{\rho}$  defined in (11) is an empirical-law-invariant coherent risk statistic.

(ii) *If  $\hat{\rho}$  is an empirical-law-invariant coherent risk statistic, then there exists a set of weights  $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^n$  such that each  $\tilde{w} \in \mathcal{W}$  satisfies (8), (9), and (10), and*

$$\hat{\rho}(\tilde{x}) = s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{j=1}^{n_1} w_j^1 x_{(j)}^1 + \sum_{j=1}^{n_2} w_j^2 x_{(j)}^2 + \cdots + \sum_{j=1}^{n_m} w_j^m x_{(j)}^m \right\}, \quad \forall \tilde{x} = (\tilde{x}^1, \dots, \tilde{x}^m) \in \mathbb{R}^n. \quad (12)$$

**PROOF.** See Appendix C.  $\square$

Theorem 3.1 and 3.3 set out the main differences between natural risk statistics and coherent risk measures: (i) Any empirical-law-invariant coherent risk statistic assigns larger weights to larger observations because both  $x_{(j)}^i$  and  $w_j^i$  increase when  $j$  increases; by contrast, natural risk statistics are more general and can assign any weights to the observations. (ii) VaR and VaR with scenario analysis, such as the Basel II and Basel III risk measures (see their definition in Section 4), are not empirical-law-invariant

coherent risk statistics because VaR does not assign larger weights to larger observations when it is estimated from data. However, VaR and VaR with scenario analysis are natural risk statistics, as will be shown in Section 4. (iii) Empirical-law-invariant coherent risk statistics are a subclass of natural risk statistics.

**3.4 Comparison with Insurance Risk Measures.** Insurance risk statistics, the data-based versions of insurance risk measures, can be defined similarly. A risk statistic  $\hat{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}$  is called an *insurance risk statistic* if it satisfies the following Axiom F1-F4.

**Axiom F1.** Empirical law 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.** Scenario-wise comonotonic additivity:  $\hat{\rho}(\tilde{x} + \tilde{y}) = \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y})$ , if  $\tilde{x}$  and  $\tilde{y}$  are scenario-wise comonotonic.

**Axiom F4.** Scale normalization:  $\hat{\rho}(\mathbf{1}) = s$ , where  $s > 0$  is a constant.

**THEOREM 3.4**  $\hat{\rho}$  is an insurance risk statistic if and only if there exists a single weight  $\tilde{w} = (w_1^1, \dots, w_{n_1}^1, \dots, w_1^m, \dots, w_{n_m}^m) \in \mathbb{R}^n$  with  $w_j^i \geq 0$  for  $j = 1, \dots, n_i; i = 1, \dots, m$  and  $\sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i = 1$ , such that

$$\hat{\rho}(\tilde{x}) = s \left[ \sum_{j=1}^{n_1} w_j^1 x_{(j)}^1 + \sum_{j=1}^{n_2} w_j^2 x_{(j)}^2 + \dots + \sum_{j=1}^{n_m} w_j^m x_{(j)}^m \right], \quad \forall \tilde{x} = (\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^m) \in \mathbb{R}^n, \quad (13)$$

where  $(x_{(1)}^i, \dots, x_{(n_i)}^i)$  is the order statistics of  $\tilde{x}^i = (x_1^i, \dots, x_{n_i}^i)$ ,  $i = 1, \dots, m$ .

**PROOF.** See Appendix D. □

Comparing Theorem 3.1 and 3.4 highlights the major differences between natural risk statistics and insurance risk measures: (i) An insurance risk statistic corresponds to a single weight vector  $\tilde{w}$ , but a natural risk statistic can incorporate multiple weights. (ii) VaR with scenario analysis, such as the Basel II and III risk measures, are not special cases of insurance risk statistics but special cases of natural risk statistics. (iii) Insurance risk statistics are a subclass of natural risk statistics.

**EXAMPLE 3.1** Although natural risk statistics include both empirical-law-invariant coherent risk statistics and insurance risk statistics, not all risk statistics are natural risk statistics. For example, the shortfall risk statistic with order  $p > 1$ , which is the data-based version of the shortfall risk measure  $\rho(X) := E[|X|^p | X > \text{VaR}_\alpha(X)]$  (see, e.g., Tasche [46]), is not a natural risk statistic. Indeed, in the one-scenario case, for a set of observations  $\tilde{x} = (x_1, \dots, x_n)$ , the shortfall risk statistic with order  $p > 1$  is defined by

$$\hat{\rho}(\tilde{x}) := \frac{1}{n - \lceil n\alpha \rceil} \sum_{k=\lceil n\alpha \rceil + 1}^n |x_{(k)}|^p.$$

Suppose that  $\tilde{x}$  and  $\tilde{y} = (y_1, \dots, y_n)$  are comonotonic, and  $x_{(k)} > 0$  and  $y_{(k)} > 0$  for all  $k > \lceil n\alpha \rceil$ , then

$$\hat{\rho}(\tilde{x} + \tilde{y}) = \frac{1}{n - \lceil n\alpha \rceil} \sum_{k=\lceil n\alpha \rceil + 1}^n (x_{(k)} + y_{(k)})^p > \frac{1}{n - \lceil n\alpha \rceil} \sum_{k=\lceil n\alpha \rceil + 1}^n (x_{(k)}^p + y_{(k)}^p) = \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y}).$$

**4. Axiomatization of the Basel II and Basel III Risk Measures.** The Basel II Accord [5] specifies that the capital charge for the trading book on any particular day  $t$  for banks using the internal models approach should be calculated by the formula

$$c_t = \max \left\{ \text{VaR}_{t-1}, k \cdot \frac{1}{60} \sum_{i=1}^{60} \text{VaR}_{t-i} \right\}, \quad (14)$$

where  $k$  is a constant that is no less than 3;  $\text{VaR}_{t-i}$  is the 10-day VaR at 99% confidence level calculated on day  $t-i$ ,  $i = 1, \dots, 60$ .  $\text{VaR}_{t-i}$  is usually estimated from a data set  $\tilde{x}^i = (x_1^i, x_2^i, \dots, x_{n_i}^i) \in \mathbb{R}^{n_i}$  which is generated by historical simulation or Monte Carlo simulation (Jorion [30]).

Adrian and Brunnermeier [2] point out that risk measures based on contemporaneous observations, such as the Basel II risk measure (14), are *procyclical*, i.e., risk measurement obtained by such risk measures tend to be low in booms and high in crises, which impedes effective regulation. Gordy and Howells [19] examine the procyclicality of Basel II from the perspective of market discipline. They show that the marginal impact of introducing Basel II depends strongly on the extent to which market discipline leads banks to vary lending standards procyclically in the absence of binding regulation. They also evaluate policy options not only in terms of their efficacy in dampening cyclicity in capital requirements, but also in terms of how well the information value of Basel II market disclosures is preserved.

Scenario analysis can help to reduce procyclicality by using not only contemporaneous observations but also data under distressed scenarios that capture rare tail events which could cause severe losses. Indeed, to respond to the financial crisis that started in late 2007, the Basel committee recently proposed the Basel III risk measure for setting capital requirements for trading books [7], which is defined by

$$c_t = \max \left\{ \text{VaR}_{t-1}, k \cdot \frac{1}{60} \sum_{i=1}^{60} \text{VaR}_{t-i} \right\} + \max \left\{ \text{sVaR}_{t-1}, \ell \cdot \frac{1}{60} \sum_{i=1}^{60} \text{sVaR}_{t-i} \right\}, \quad (15)$$

where  $\text{VaR}_{t-i}$  is the same as in (14);  $k$  and  $\ell$  are constants no less than 3;  $\text{sVaR}_{t-i}$  is called the *stressed* VaR on day  $t-i$ , which is calculated under the scenario that the financial market is under significant stress as happened during the period from 2007 to 2008. The additional capital requirements based on stressed VaR help reduce the procyclicality of the original risk measure (14).

In addition to the capital charge specified in (15), the Basel III Accord requires banks to hold additional incremental risk capital charge (IRC) against potential losses resulting from default risk, credit migration risk, credit spread risk, etc. in the trading book which are incremental to the risks captured by the formula (15) [6, 7]. The IRC capital charge on the  $t$ -th day is defined as

$$\text{IRC}_t = \max \left\{ \text{VaR}_{t-1}^{\text{ir}}, \frac{1}{60} \sum_{i=1}^{60} \text{VaR}_{t-i}^{\text{ir}} \right\}, \quad (16)$$

where  $\text{VaR}_{t-i}^{\text{ir}}$  is defined as the 99.9% VaR of the trading book loss due to the aforementioned risks over a one-year horizon calculated on day  $t-i$ . The  $\text{VaR}_{t-i}^{\text{ir}}$  should be calculated under the assumption that the portfolio is rebalanced to maintain a target level of risk and that less liquid assets have long liquidity horizons (see [6]). Glasserman [17] analyzes the features of the IRC risk measure, with particular emphasis on the impact of the liquidity horizons nested within the long risk horizon of one year on the portfolio's loss distribution.

The Basel II and Basel III risk measures do not belong to any existing theoretical framework of risk measures proposed in the literature, but they are special cases of natural risk statistics, as is shown by the following theorems.

**THEOREM 4.1** *The Basel II risk measure defined in (14) and the Basel III risk measure defined in (15) are both special cases of natural risk statistics.*

**PROOF.** See Appendix E. □

**THEOREM 4.2** *The Basel III risk measure for incremental risk defined in (16) is a special case of natural risk statistics.*

**PROOF.** See Appendix E. □

Natural risk statistics thus provide an axiomatic framework for understanding and, if necessary, extending the Basel accords. Having such a general framework then facilitates searching for other external risk measures suitable for banking regulation.

**EXAMPLE 4.1** *The regulators may have different objectives in choosing external risk measures. For example, as we shall explain in the next section, it is desirable to make them robust. Another objective is to choose less pro-cyclical risk measures. Gordy and Howells [19] propose to mitigate the procyclicality of  $c_t$ , the Basel II capital requirement, by a method called counter-cyclical indexing. This applies a time-varying*



multiplier  $\alpha_t$  to  $c_t$  and generates a smoothed capital requirement  $\alpha_t c_t$ , where  $\alpha_t$  increases during booms and decreases during recessions to dampen the procyclicality of  $c_t$ . In the static setting the multiplier  $\alpha_t$  corresponds to the scaling constant  $s$  in Axiom C1; thus, natural risk statistics provide an axiomatic foundation in the static setting for the method of counter-cyclical indexing. Although the current paper focuses on static risk measures, it would be of interest to study axioms for dynamic risk measures which also depend on business cycles.

## 5. Robustness of External Risk Measures.

**5.1 The Meaning of Robustness.** A risk measure is said to be *robust* if (i) it can accommodate model misspecification (possibly by incorporating multiple scenarios and models); and (ii) it is insensitive to small changes in the data, i.e., small changes in all, or large changes in a few of the samples (possibly by using robust statistics).

The first part of the meaning of robustness is related to ambiguity and model uncertainty in decision theory. To address these issues, multiple priors or multiple alternative models represented by a set of probability measures may be used; see, e.g., Gilboa and Schmeidler [16], Maccheroni, Marinacci, and Rustichini [35], and Hansen and Sargent [20]. The second part of the meaning of robustness comes from the study of robust statistics, which is mainly concerned with the statistical distribution robustness; see, e.g., Huber and Ronchetti [25]. Appendix F presents a detailed mathematical discussion of robustness.

**5.2 Legal Background.** Legal realism, one of the basic concepts of law, motivates us to argue that external risk measures should be robust because robustness is essential for law enforcement. Legal realism is the viewpoint that the legal decisions of a court are determined by the actual practices of the judges rather than the law set forth in statutes and precedents. All the legal rules contained in statutes and precedents are uncertain due to the uncertainty in human language and the fact that human beings are unable to anticipate all possible future circumstances (Hart [21, p. 128]). Hence, a law is only a guideline for judges and enforcement officers (Hart [21, pp. 204–205]), i.e., it 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 it. In particular, consistent enforcement of an external risk measure in banking regulation requires that it should be robust with respect to underlying models and data.

An illuminating example manifesting the concept of legal realism is how to set up speed limits on roads, which is a crucial issue involving life and death decisions. Currently, American Association of State Highway and Transportation Officials recommends setting speed limits near the 85th percentile of the free flowing traffic speed observed on the road with an adjustment taking into consideration that people tend to drive 5 to 10 miles above the posted speed limit (Transportation Research Board of the National Academies [47, p. 51]). This recommendation is adopted by all states and most local agencies. The 85th percentile rule appears to be a simple method, but studies have shown that crash rates are lowest at around the 85th percentile. The 85th percentile rule is robust in the sense that it is based on data rather than on some subjective model and it can be implemented consistently.

**5.3 Robustness Is Indispensable for External Risk Measures.** In determining capital requirements, regulators impose a risk measure and allow institutions to use their own internal risk models and private data in the calculation. For example, the internal model approach in Basel II and III allows institutions to use their own internal models to calculate their capital requirements for trading books due to various legal, commercial, and proprietary trading considerations. However, there are two issues arising from the use of internal models and private data in external regulation: (i) the data can be noisy, flawed, or unreliable; and (ii) there can be several statistically indistinguishable models for the same asset or portfolio due to limited availability of data. For example, the heaviness of tail distributions cannot be identified in many cases. Heyde and Kou [22] 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) because the quantiles of exponential-type distributions and power-type distributions may overlap. For example, surprisingly, a Laplace distribution has a *larger* 99.9% quantile than the corresponding T distribution with degree of freedom (d.f.) 6 or 7. Hence, regardless of the sample size, the Laplace distribution may appear to be more heavily tailed than the T distribution up to the 99.9% quantile. If the quantiles

have to be estimated from data, the situation is even worse. In fact, with a sample size of 5,000 it is difficult to distinguish between the Laplace distribution and the T distributions with d.f. 3, 4, 5, 6, and 7 because the asymptotic 95% confidence interval of the 99.9% quantile of the Laplace distribution overlaps with those of the T distributions. Therefore, the tail behavior may be a subjective issue depending on people's modeling preferences.

To address the aforementioned two issues, external risk measures should demonstrate robustness with respect to model misspecification and small changes in the data. From a regulator's viewpoint, an external risk measure must be unambiguous, stable, and capable of being implemented consistently across all the relevant institutions, no matter what internal beliefs or internal models each may rely on. When the correct model cannot be identified, two institutions that have exactly the same portfolio can use different internal models, both of which can obtain the approval of the regulator; however, the two institutions should be required to hold the same or at least almost the same amount of regulatory capital because they have the same portfolio. Therefore, the external risk measure should be robust; otherwise, different institutions can be required to hold very different regulatory capital for the same risk exposure, which makes the risk measure unacceptable to both the institutions and the regulators. In addition, if the external risk measure is not robust, institutions can take regulatory arbitrage by choosing a model that significantly reduces the capital requirements or by manipulating the input data.

**5.4 Tail Conditional Median: a Robust Risk Measure.** We propose a robust risk measure, *tail conditional median* (TCM), which is a special case of natural risk statistics. TCM of the random loss  $X$  at level  $\alpha$  is defined as

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

For  $X$  with a continuous distribution, it holds that

$$\text{TCM}_\alpha(X) = \text{median}[X|X \geq \text{VaR}_\alpha(X)] = \text{VaR}_{\frac{1+\alpha}{2}}(X). \quad (17)$$

For  $X$  with a general distribution having discontinuities,  $\text{VaR}_{\frac{1+\alpha}{2}}(X)$  may be slightly different from  $\text{TCM}_\alpha(X)$  if either  $\text{VaR}_{\frac{1+\alpha}{2}}(X)$  or  $\text{TCM}_\alpha(X)$  happens to be equal to a discontinuity.

Eq. (17) shows that VaR at a higher level can incorporate tail information, which contradicts the claims in some of the existing literature. For example, if one wants to measure the size of loss beyond the 99% level, one can use VaR at 99.5%, or equivalently TCM at 99%, which gives the median of the size of loss beyond 99%. It is also interesting to point out that  $\text{TCM}_\alpha(X + Y) \leq \text{TCM}_\alpha(X) + \text{TCM}_\alpha(Y)$  may hold for those  $X$  and  $Y$  that cause  $\text{VaR}_\alpha(X + Y) > \text{VaR}_\alpha(X) + \text{VaR}_\alpha(Y)$ ; in other words, subadditivity may not be violated if one replaces  $\text{VaR}_\alpha$  by  $\text{TCM}_\alpha$ . Here are two such examples: (i) The example on page 216 of Artzner et al. [4] shows that 99% VaR does not satisfy subadditivity for the two positions of writing an option  $A$  and writing an option  $B$ . However, 99% TCM (or equivalently 99.5% VaR) does satisfy subadditivity. Indeed, the 99% TCM of the three positions of writing an option  $A$ , writing an option  $B$ , and writing options  $A + B$  are equal to  $1000 - u$ ,  $1000 - l$ , and  $1000 - u - l$ , respectively. (ii) The example on page 217 of Artzner et al. [4] shows that the 90% VaR does not satisfy subadditivity for  $X_1$  and  $X_2$ . However, the 90% TCM (or equivalently 95% VaR) does satisfy subadditivity. Actually, the 90% TCM 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 90% TCM of  $X_1 + X_2$  is strictly less than 2.

TCM can be shown to be more robust than TCE by at least four tools in robust statistics: (i) influence functions, (ii) asymptotic breakdown points, (iii) continuity of statistical functional, and (iv) finite sample breakdown points. See Appendix F.

TCE is also highly model-dependent, and particularly sensitive to modeling assumptions on the extreme tails of loss distributions, because the computation of TCE relies on these extreme tails, as is shown by (66) in Appendix F. Figure 1 illustrates the sensitivity of TCE to modeling assumptions. TCM is clearly less sensitive to tail behavior than TCE, since the changes in TCM have narrower ranges than those in TCE.

**5.5 Robust Natural Risk Statistics.** Natural risk statistics include a subclass of risk statistics that are robust in two respects: (i) they are insensitive to model misspecification because they incorporate multiple scenarios, multiple prior probability measures on the set of scenarios, and multiple subsidiary

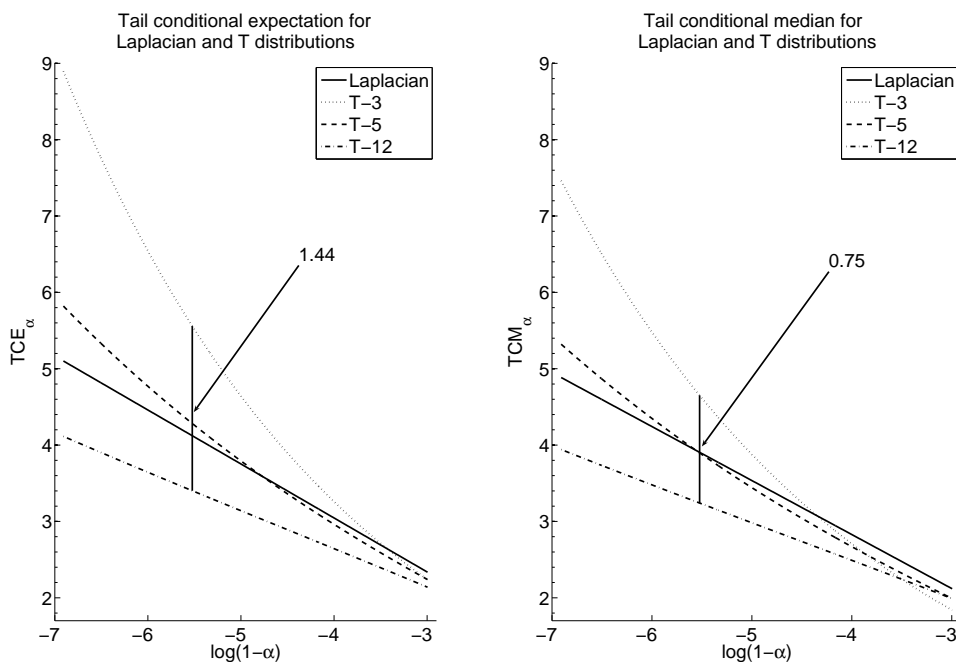


Figure 1: Comparing the robustness of tail conditional expectation (TCE) and tail conditional median (TCM) with respect to model misspecification.  $TCE_\alpha$  and  $TCM_\alpha$  are calculated for Laplace and T distributions with degree of freedom 3, 5, 12, which are normalized to have mean 0 and variance 1. The horizontal axis is  $\log(1 - \alpha)$  for  $\alpha \in [0.95, 0.999]$ . For  $\alpha = 99.6\%$ , the variation of  $TCE_\alpha$  with respect to the change in the underlying models is 1.44, but the variation of  $TCM_\alpha$  is only 0.75.

risk statistics for each scenario; and (ii) they are insensitive to small changes in the data because they use robust statistics for each scenario.

Let  $\hat{\rho}$  be a natural risk statistic defined as in (5) that corresponds to the set of weights  $\mathcal{W}$ . Define the map  $\phi : \mathcal{W} \rightarrow \mathbb{R}^m \times \mathbb{R}^n$  such that  $\tilde{w} \mapsto \phi(\tilde{w}) := (\tilde{p}, \tilde{q})$ , where  $\tilde{p} := (p^1, \dots, p^m)$ ,  $p^i := \sum_{j=1}^{n_i} w_j^i$ ,  $i = 1, \dots, m$ ;  $\tilde{q} := (q_1^1, \dots, q_{n_1}^1, \dots, q_1^m, \dots, q_{n_m}^m)$ ,  $q_j^i := 1_{\{p_i > 0\}} w_j^i / p_i$ . Since  $p^i \geq 0$  and  $\sum_{i=1}^m p^i = 1$ ,  $\tilde{p}$  can be viewed as a prior probability distribution on the set of scenarios. Then  $\hat{\rho}$  can be rewritten as

$$\hat{\rho}(\tilde{x}) = s \cdot \sup_{(\tilde{p}, \tilde{q}) \in \phi(\mathcal{W})} \left\{ \sum_{i=1}^m p^i \hat{\rho}^{i, \tilde{q}}(\tilde{x}^i) \right\}, \quad \text{where } \hat{\rho}^{i, \tilde{q}}(\tilde{x}^i) := \sum_{j=1}^{n_i} q_j^i x_{(j)}^i. \quad (18)$$

Each weight  $\tilde{w} \in \mathcal{W}$  then corresponds to  $\phi(\tilde{w}) = (\tilde{p}, \tilde{q}) \in \phi(\mathcal{W})$ , which specifies: (i) the prior probability measure  $\tilde{p}$  on the set of scenarios; and (ii) the subsidiary risk statistic  $\hat{\rho}^{i, \tilde{q}}$  for each scenario  $i$ ,  $i = 1, \dots, m$ . Hence,  $\hat{\rho}$  can be robust with respect to model misspecification by incorporating multiple prior probabilities  $\tilde{p}$  and multiple risk statistics  $\hat{\rho}^{i, \tilde{q}}$  for each scenario. In addition,  $\hat{\rho}$  can be robust with respect to small changes in the data if each subsidiary risk statistic  $\hat{\rho}^{i, \tilde{q}}$  is a robust L-statistic.

EXAMPLE 5.1 *Examples of robust L-statistics (see Huber and Ronchetti [25, Chap. 3]) include TCM (or equivalently VaR at a higher confidence level) and the following risk statistic:*

$$\hat{\rho}^{i, \tilde{q}}(\tilde{x}^i) := \frac{1}{[\beta n_i] - [\alpha n_i]} \sum_{j=[\alpha n_i]+1}^{[\beta n_i]} x_{(j)}^i, \quad (19)$$

where  $0 < \alpha < \beta < 1$ . The above defined statistic is an estimate of the “Truncated TCE”  $\rho(X) := E[X | VaR_\alpha(X) < X \leq VaR_\beta(X)]$  using samples observed under the  $i$ -th scenario.

EXAMPLE 5.2 *The Basel II risk measure (14) is robust to a certain extent because (i) each subsidiary risk statistic is a VaR, which is robust; and (ii) the risk measure incorporates two priors of probability distributions on the set of scenarios. More precisely, one prior assigns probability  $\frac{1}{k}$  to the scenario of*

day  $t-1$  and  $1 - \frac{1}{k}$  to an imaginary scenario under which losses are identically 0; the other prior assigns probability  $\frac{1}{60}$  to each of the scenarios corresponding to day  $t-i$ ,  $i = 1, \dots, 60$ .

**EXAMPLE 5.3** *The Basel III risk measure (15) is more robust than the Basel II risk measure (14) because it incorporates 60 more scenarios and it essentially incorporates two more priors of probability measures on the set of scenarios.*

**EXAMPLE 5.4** *Similar to the Basel II risk measure (14), the Basel III IRC risk measure (16) is robust in the sense that each subsidiary risk statistic  $\text{VaR}_{t-i}^{\text{IR}}$  is robust, and the risk measure incorporates two priors of probability distributions on the set of scenarios.*

**5.6 Neither Coherent Risk Measures Nor Insurance Risk Measures Are Robust.** No coherent risk measure is robust with respect to small changes in the data because coherent risk measures exclude the use of robust statistics. Indeed, by Theorem 3.3, an empirical-law-invariant coherent risk statistic  $\hat{\rho}$  can be represented by (12), where for each weight  $\tilde{w}$ ,  $w_j^i$  is a nondecreasing function of  $j$ . Hence, any empirical-law-invariant coherent risk statistic assigns larger weights to larger observations, but assigning larger weights to larger observations is clearly sensitive to small changes in the data. An extreme case is the maximum loss  $\max\{x_{(n_i)}^i : i = 1, \dots, m\}$ , which is not robust at all. In general, the finite sample breakdown point (see Huber and Ronchetti [25, Chap. 11] for definition) of any empirical-law-invariant coherent risk statistic is equal to  $1/(1+n)$ , which implies that one single contamination sample can cause unbounded bias. In particular, TCE is sensitive to modeling assumptions of heaviness of tail distributions and to outliers in the data, as is shown in Section 5.4.

No insurance risk measure is robust to model misspecification. An insurance risk measure can incorporate neither multiple priors of probability distributions on the set of scenarios nor multiple subsidiary risk statistics for each scenario because it is defined by a single weight vector  $\tilde{w}$ , as is shown in Theorem 3.4.

**5.7 Conservative and Robust Risk Measures.** One risk measure is said to be more conservative than another if it generates higher risk measurement than the other for the same 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 which are too conservative may retard 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 a good external risk measure. Conservativeness addresses the issue of how stringently an external risk measure should be implemented, given that it can be implemented consistently. In other words, 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. In addition, it is not true that TCE is more conservative than TCM because the median can be bigger than the mean for some distributions. Eling and Tibiletti [12] 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. They find that although TCE is on average about 10% higher than TCM at standard confidence levels, TCM is higher than TCE in about 10% of the cases. So TCE is not necessarily more conservative than TCM.

A natural risk statistic can be constructed by (5) in the following way so that it is both conservative and robust: (i) more data subsets that correspond to stressed scenarios can be included in (5); and (ii) a larger constant  $s$  in (5) can be used. For example, adding 60 stressed scenarios makes (15) much more conservative than (14), and a larger  $k$  or  $\ell$  in (15) can be used by regulators to increase the capital requirements.

## 6. Other Reasons to Relax Subadditivity.

**6.1 Diversification and Tail Subadditivity of VaR.** The subadditivity axiom is related to the idea that diversification does not increase risk; the convexity axiom for convex risk measures also comes from the idea of diversification. There are two main justifications for diversification. One is based on the simple observation that  $\sigma(X+Y) \leq \sigma(X) + \sigma(Y)$ , for any two random variables  $X$  and  $Y$  with

finite second moments, where  $\sigma(\cdot)$  denotes standard deviation. The other is based on expected utility theory. Samuelson [39] 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; see, e.g., McMinn [36], Hong and Herk [23], and Kijima [32] for the discussion on whether diversification is beneficial when the asset returns are dependent. Both justifications require that the risks have finite second moments.

Is diversification still preferable for risks with infinite second moments? The answer can be no. Ibragimov [26, 27] and Ibragimov and Walden [28] show that diversification is not preferable for risks with extremely heavy tailed distributions (with tail index less than 1) in the sense that: (i) the loss of the diversified portfolio stochastically dominates that of the undiversified portfolio at the first order and second order; (ii) the expected utility of the (truncated) payoff 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.

In addition, the conclusion that VaR prohibits diversification, drawn from simple examples in the literature, may not be solid. For instance, Artzner et al. [4] show that VaR prohibits diversification by a simple example (see pp. 217–218) in which 95% VaR of the diversified portfolio is higher than that of the undiversified portfolio. However, in the same example 99% VaR encourages diversification because the 99% VaR of the diversified portfolio is equal to 20,800, which is much lower than 1,000,000, the 99% VaR of the undiversified portfolio.

Ibragimov [26, 27] and Ibragimov and Walden [28] also show that although VaR does not satisfy subadditivity for risks with extremely heavy tailed distributions (with tail index less than 1), VaR satisfies subadditivity for wide classes of independent and dependent risks with tail indices greater than 1. In addition, Daniélsson, Jorgensen, Samorodnitsky, Sarma, and de Vries [8] show that VaR is subadditive in the tail region provided that the tail index of the joint distribution is larger than 1. Asset returns with tail indices less than 1 have extremely heavy tails; they are hard to find but easy to identify. Daniélsson et al. [8] argue that they can be treated as special cases in financial modeling. Even if one encounters an extremely fat tail and insists on tail subadditivity, Garcia, Renault, and Tsafack [15] 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. The simulations carried out in Daniélsson et al. [8] also show that  $\text{VaR}_\alpha$  is indeed subadditive for most practical applications when  $\alpha \in [95\%, 99\%]$ .

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

**6.2 Does A Merger Always Reduce Risk?** Subadditivity basically means that “a merger does not create extra risk” (see Artzner et al. [4, p. 209]). However, Dhaene, Goovaerts, and Kaas [11] point out that a merger may increase risk, particularly when there is bankruptcy protection for institutions. For example, an institution can split a risky trading business into a separate subsidiary so that it has the option to let the subsidiary go bankrupt when the subsidiary suffers enormous losses, confining losses to that subsidiary. Therefore, creating subsidiaries may incur less risk and a merger may increase risk. Had Barings Bank set up a separate institution for its Singapore unit, the bankruptcy of that unit would not have sunk the entire bank in 1995.

In addition, there is little empirical evidence supporting the argument that “a merger does not create extra risk.” In practice, credit rating agencies do not upgrade an institution’s credit rating because of a merger; on the contrary, the credit rating of the joint institution may be lowered shortly after the merger. The merger of Bank of America and Merrill Lynch in 2008 is an example.

**7. Capital Allocation under the Natural Risk Statistics.** In this section, we derive the capital allocation rule for a subclass of natural risk statistics which include the Basel II and Basel III risk measures. The purpose of capital allocation for the whole portfolio is to decompose the overall capital into a sum of risk contributions for such purposes as identification of concentration, risk-sensitive pricing, and portfolio optimization (see, e.g., Litterman [34]).

First, as an illustration, we compute the Euler capital allocation under the Basel III risk measure.

The Euler rule is one of the most widely used methodologies for capital allocation under positive homogeneous risk measures (see, e.g., Tasche [46]; McNeil, Frey, and Embrechts [37]). Consider a portfolio comprised of  $u_i$  units of asset  $i$ ,  $i = 1, \dots, d$ , and denote  $u = (u_1, u_2, \dots, u_d)$ . Suppose that there are  $m$  scenarios. Let  $\tilde{x}(i) = (\tilde{x}(i)^1, \tilde{x}(i)^2, \dots, \tilde{x}(i)^m)$  be the observed loss of the  $i$ -th asset, where  $\tilde{x}(i)^s = (x(i)_1^s, x(i)_2^s, \dots, x(i)_{n_s}^s) \in \mathbb{R}^{n_s}$  are the observations under the  $s$ -th scenario,  $s = 1, \dots, m$ . Then the observations of the portfolio loss are given by  $\tilde{l}(u) = \sum_{i=1}^d u_i \tilde{x}(i) = (\tilde{l}(u)^1, \tilde{l}(u)^2, \dots, \tilde{l}(u)^m)$ , where  $\tilde{l}(u)^s = (l(u)_1^s, l(u)_2^s, \dots, l(u)_{n_s}^s) \in \mathbb{R}^{n_s}$  and  $l(u)_k^s := \sum_{i=1}^d u_i x(i)_k^s$ . The required capital measured by a natural risk statistic  $\hat{\rho}$  is denoted by  $C_{\hat{\rho}}(u) := \hat{\rho}(\tilde{l}(u))$ . Let  $m = 120$  and  $\alpha = 99\%$ , then the required capital calculated by the Basel III risk measure is

$$C_{\hat{\rho}}(u) := \max \left\{ l(u)_{(\lceil \alpha n_1 \rceil)}^1, \frac{k}{60} \sum_{s=1}^{60} l(u)_{(\lceil \alpha n_s \rceil)}^s \right\} + \max \left\{ l(u)_{(\lceil \alpha n_{61} \rceil)}^{61}, \frac{\ell}{60} \sum_{s=61}^{120} l(u)_{(\lceil \alpha n_s \rceil)}^s \right\}.$$

We have the following proposition on the Euler capital allocation under the Basel III risk statistic:

**PROPOSITION 7.1** *Suppose  $\tilde{x}$  is a sample of the random vector  $(X(1), X(2), \dots, X(d))$ , where  $X(i) = (X(i)^1, X(i)^2, \dots, X(i)^m)$  and  $X(i)^s = (X(i)_1^s, X(i)_2^s, \dots, X(i)_{n_s}^s) \in \mathbb{R}^{n_s}$ . Suppose that the joint distribution of  $(X(1), X(2), \dots, X(d))$  has a probability density on  $\mathbb{R}^{dn}$ . Then for any given  $u \neq 0$ , it holds with probability 1 that*

$$C_{\hat{\rho}}(u) = \sum_{i=1}^d u_i \frac{\partial C_{\hat{\rho}}(u)}{\partial u_i}, \quad (20)$$

and the capital allocation for the  $i$ -th asset under the Euler's rule is  $u_i \frac{\partial C_{\hat{\rho}}(u)}{\partial u_i}$ .

**PROOF.** For any given  $u \neq 0$ , let  $\mathbb{X}_u$  be the set of samples  $(\tilde{x}(1), \tilde{x}(2), \dots, \tilde{x}(d)) \in \mathbb{R}^{dn}$  that satisfy the following conditions: (i)  $l(u)_{(\lceil \alpha n_1 \rceil)}^1 \neq \frac{k}{60} \sum_{s=1}^{60} l(u)_{(\lceil \alpha n_s \rceil)}^s$ ; (ii)  $l(u)_{(\lceil \alpha n_{61} \rceil)}^{61} \neq \frac{\ell}{60} \sum_{s=61}^{120} l(u)_{(\lceil \alpha n_s \rceil)}^s$ ; (iii)  $l(u)_i^s \neq l(u)_j^s$  for any  $s$  and  $i \neq j$ . Then it follows from the condition of the proposition that  $P((X(1), X(2), \dots, X(d)) \in \mathbb{X}_u) = 1$ . Fix any  $(\tilde{x}(1), \tilde{x}(2), \dots, \tilde{x}(d)) \in \mathbb{X}_u$ . By the definition of  $\mathbb{X}_u$ , there exists  $\delta > 0$ , such that  $C_{\hat{\rho}}(\cdot)$  is a linear function on the open set  $V := \{v \in \mathbb{R}^d : \|v - u\| < \delta\}$ . Hence,  $C_{\hat{\rho}}(\cdot)$  is differentiable at  $u$  and Eq. (20) holds.  $\square$

Let  $\mathbb{X}_u$  be defined in the above proof and  $\tilde{x} \in \mathbb{X}_u$ . To compute  $u_i \frac{\partial C_{\hat{\rho}}(u)}{\partial u_i}$ , one only needs to compute  $\frac{\partial l(u)_{(\lceil \alpha n_s \rceil)}^s}{\partial u_i}$ . Let  $(p_1, \dots, p_{n_s})$  be the permutation of  $(1, 2, \dots, n_s)$  such that  $l(u)_{p_1}^s < l(u)_{p_2}^s < \dots < l(u)_{p_{n_s}}^s$ . Then there exists a neighborhood  $V := \{v \in \mathbb{R}^d : \|v - u\| < \delta\}$  of  $u$  such that  $l(v)_{p_1}^s < l(v)_{p_2}^s < \dots < l(v)_{p_{n_s}}^s$  for  $\forall v \in V$ . Hence, for  $\forall v \in V$ ,  $l(v)_{(\lceil \alpha n_s \rceil)}^s = l(v)_{p_{\lceil \alpha n_s \rceil}}^s = \sum_{i=1}^d v_i x(i)_{p_{\lceil \alpha n_s \rceil}}^s$ , and  $\frac{\partial l(u)_{(\lceil \alpha n_s \rceil)}^s}{\partial u_i} = x(i)_{p_{\lceil \alpha n_s \rceil}}^s$ .

In general, let  $\Upsilon_1$  be the set of natural risk statistic  $\hat{\rho}$  that can be represented in Eq. (6) by a finite set  $\mathcal{W}$ . Let  $\Upsilon_2$  be the set of natural risk statistic  $\hat{\rho}$  that can be written as  $\hat{\rho} = \sum_{k=1}^K a_k \hat{\rho}_k$ , where  $a_k \geq 0$  and  $\hat{\rho}_k \in \Upsilon_1$ ,  $k = 1, \dots, K$ . Both the Basel II risk measure and Basel III risk measure belong to the set  $\Upsilon_2$ . For any  $\hat{\rho} \in \Upsilon_2$ , it can be shown in the same way as in Proposition 7.1 that  $C_{\hat{\rho}}(u)$  is a piece-wise linear function of  $u$  and the Euler capital allocation rule can be computed similarly.

**8. Conclusion.** We propose a class of 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 required by coherent risk measures, and relaxing the comonotonic additivity required by insurance risk measures.

Natural risk statistics include VaR with scenario analysis, and particularly the Basel II and Basel III risk measures, as special cases. Thus, natural risk statistics provide a theoretical framework for understanding and, if necessary, extending the Basel accords. Indeed, the framework is general enough to include the counter-cyclical indexing risk measure suggested by Gordy and Howells [19] to address the procyclicality problem in Basel II.

We emphasize that an external risk measure should be robust to model misspecification and small changes in the data in order for its consistent implementation across different institutions. We show that

coherent risk measures are generally not robust with respect to small changes in the data and insurance risk measures are generally not robust with respect to model misspecification.

Natural risk statistics include a subclass of robust risk measures that are suitable for external regulation. In particular, natural risk statistics include tail conditional median (with scenario analysis), which is more robust than tail conditional expectation suggested by the theory of coherent risk measures. The Euler capital allocation can also be easily calculated under the natural risk statistics.

**Appendix A. Proof of Theorem 3.1.** A simple observation is that  $\hat{\rho}$  is a natural risk statistic corresponding to a constant  $s$  in Axiom C1 if and only if  $\frac{1}{s}\hat{\rho}$  is a natural risk statistic corresponding to the constant  $s = 1$  in Axiom C1. Therefore, in this section, we assume without loss of generality that  $s = 1$  in Axiom C1. The proof relies on the following two lemmas, which depend heavily on the properties of the interior points of the set

$$\mathcal{B} := \{\tilde{y} = (\tilde{y}^1, \dots, \tilde{y}^m) \in \mathbb{R}^n \mid y_1^1 \leq y_2^1 \leq \dots \leq y_{n_1}^1; \dots; y_1^m \leq y_2^m \leq \dots \leq y_{n_m}^m\}. \quad (21)$$

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 A.1** *Let  $\mathcal{B}$  be defined in (21) and  $\mathcal{B}^\circ$  be the interior of  $\mathcal{B}$ . For any fixed  $\tilde{z} \in \mathcal{B}^\circ$  and any  $\hat{\rho}$  satisfying Axiom C1–C4 and  $\hat{\rho}(\tilde{z}) = 1$ , there exists a weight  $\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^m) \in \mathbb{R}^n$  such that the linear functional  $\lambda(\tilde{x}) := \sum_{j=1}^{n_1} w_j^1 x_j^1 + \sum_{j=1}^{n_2} w_j^2 x_j^2 + \dots + \sum_{j=1}^{n_m} w_j^m x_j^m$  satisfies*

$$\lambda(\tilde{z}) = 1, \quad (22)$$

$$\lambda(\tilde{x}) < 1 \text{ for any } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x}) < 1. \quad (23)$$

**PROOF.** Let  $U = \{\tilde{x} = (\tilde{x}^1, \dots, \tilde{x}^m) \mid \hat{\rho}(\tilde{x}) < 1\} \cap \mathcal{B}$ . For any  $\tilde{x} = (\tilde{x}^1, \dots, \tilde{x}^m) \in \mathcal{B}$  and  $\tilde{y} = (\tilde{y}^1, \dots, \tilde{y}^m) \in \mathcal{B}$ ,  $\tilde{x}$  and  $\tilde{y}$  are scenario-wise comonotonic. Then Axiom C1 and C3 imply that  $U$  is convex, and, hence, 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  $\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^m) = (w_1^1, \dots, w_{n_1}^1, \dots, w_1^m, \dots, w_{n_m}^m) \in \mathbb{R}^n$  such that  $\lambda(\tilde{x}) := \sum_{j=1}^{n_1} w_j^1 x_j^1 + \sum_{j=1}^{n_2} w_j^2 x_j^2 + \dots + \sum_{j=1}^{n_m} w_j^m x_j^m$  satisfies  $\lambda(\tilde{x}) \leq \lambda(\tilde{z})$  for any  $\tilde{x} \in \bar{U}$ . In particular, we have

$$\lambda(\tilde{x}) \leq \lambda(\tilde{z}), \forall \tilde{x} \in U. \quad (24)$$

We shall show that the strict inequality holds in (24). Suppose, by contradiction, that there exists  $\tilde{r} \in U$  such that  $\lambda(\tilde{r}) = \lambda(\tilde{z})$ . For any  $\alpha \in (0, 1)$ , we have

$$\lambda(\alpha \tilde{z} + (1 - \alpha)\tilde{r}) = \alpha \lambda(\tilde{z}) + (1 - \alpha)\lambda(\tilde{r}) = \lambda(\tilde{z}). \quad (25)$$

In addition, since  $\tilde{z}$  and  $\tilde{r}$  are scenario-wise comonotonic, we have

$$\hat{\rho}(\alpha \tilde{z} + (1 - \alpha)\tilde{r}) \leq \alpha \hat{\rho}(\tilde{z}) + (1 - \alpha)\hat{\rho}(\tilde{r}) < \alpha + (1 - \alpha) = 1, \quad \forall \alpha \in (0, 1). \quad (26)$$

Since  $\tilde{z} \in \mathcal{B}^\circ$ , it follows that there exists  $\alpha_0 \in (0, 1)$  such that  $\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r} \in \mathcal{B}^\circ$ . Hence, for any small enough  $\varepsilon > 0$ ,

$$\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r} + \varepsilon \tilde{w} \in \mathcal{B}. \quad (27)$$

With  $w_{\max} := \max\{w_1^1, w_2^1, \dots, w_{n_1}^1; w_1^2, w_2^2, \dots, w_{n_2}^2; \dots; w_1^m, w_2^m, \dots, w_{n_m}^m\}$ , we have  $\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r} + \varepsilon \tilde{w} \leq \alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r} + \varepsilon w_{\max} \mathbf{1}$ . Thus, the monotonicity in Axiom C2 and translation scaling in Axiom C1 yield

$$\hat{\rho}(\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r} + \varepsilon \tilde{w}) \leq \hat{\rho}(\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r} + \varepsilon w_{\max} \mathbf{1}) = \hat{\rho}(\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r}) + \varepsilon w_{\max}. \quad (28)$$

Since  $\hat{\rho}(\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r}) < 1$  via (26), we have by (28) and (27) that for any small enough  $\varepsilon > 0$ ,  $\hat{\rho}(\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r} + \varepsilon \tilde{w}) < 1$ ,  $\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r} + \varepsilon \tilde{w} \in U$ . Hence, (24) implies  $\lambda(\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r} + \varepsilon \tilde{w}) \leq \lambda(\tilde{z})$ . However, we have, by (25), an opposite inequality  $\lambda(\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r} + \varepsilon \tilde{w}) = \lambda(\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r}) + \varepsilon |\tilde{w}|^2 > \lambda(\alpha_0 \tilde{z} + (1 - \alpha_0)\tilde{r}) = \lambda(\tilde{z})$ , leading to a contradiction. In summary, we have shown that

$$\lambda(\tilde{x}) < \lambda(\tilde{z}), \forall \tilde{x} \in U. \quad (29)$$

Since  $\hat{\rho}(0) = 0$ , we have  $0 \in U$ . Letting  $\tilde{x} = 0$  in (29) yields  $\lambda(\tilde{z}) > 0$ , so we can re-scale  $\tilde{w}$  such that  $\lambda(\tilde{z}) = 1 = \hat{\rho}(\tilde{z})$ . Thus, (29) becomes  $\lambda(\tilde{x}) < 1$  for any  $\tilde{x}$  such that  $\tilde{x} \in \mathcal{B}$  and  $\hat{\rho}(\tilde{x}) < 1$ , from which (23) holds.  $\square$

LEMMA A.2 Let  $\mathcal{B}$  be defined in (21) and  $\mathcal{B}^\circ$  be the interior of  $\mathcal{B}$ . For any fixed  $\tilde{z} \in \mathcal{B}^\circ$  and any  $\hat{\rho}$  satisfying Axiom C1-C4, there exists a weight  $\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^m) \in \mathbb{R}^n$  such that  $\tilde{w}$  satisfies (3) and (4), and

$$\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i \text{ for any } \tilde{x} \in \mathcal{B}, \text{ and } \hat{\rho}(\tilde{z}) = \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i z_j^i. \quad (30)$$

PROOF. We will show this by considering three cases.

Case 1:  $\hat{\rho}(\tilde{z}) = 1$ . From Lemma A.1, there exists a weight  $\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^m) \in \mathbb{R}^n$  such that the linear functional  $\lambda(\tilde{x}) := \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i$  satisfies (22) and (23).

Firstly, we prove that  $\tilde{w}$  satisfies (3), which is equivalent to  $\lambda(\mathbf{1}) = 1$ . To this end, first note that for any  $c < 1$ , Axiom C1 implies  $\hat{\rho}(c\mathbf{1}) = c < 1$ . Thus, (23) implies  $\lambda(c\mathbf{1}) < 1$ , and, by continuity of  $\lambda(\cdot)$ , we obtain that  $\lambda(\mathbf{1}) \leq 1$ . On the other hand, 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 (23) and (22) 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 (3).

Secondly, we prove that  $\tilde{w}$  satisfies (4). For any fixed  $i$  and  $1 \leq j \leq n_i$ , let  $k = n_1 + n_2 + \dots + n_{i-1} + j$  and  $\tilde{e} = (0, \dots, 0, 1, 0, \dots, 0)$  be the  $k$ -th standard basis of  $\mathbb{R}^n$ . Then  $w_j^i = \lambda(\tilde{e})$ . Since  $\tilde{z} \in \mathcal{B}^\circ$ , there exists  $\delta > 0$  such that  $\tilde{z} - \delta\tilde{e} \in \mathcal{B}$ . For any  $\varepsilon > 0$ , Axiom C1 and C2 imply  $\hat{\rho}(\tilde{z} - \delta\tilde{e} - \varepsilon\mathbf{1}) = \hat{\rho}(\tilde{z} - \delta\tilde{e}) - \varepsilon \leq \hat{\rho}(\tilde{z}) - \varepsilon = 1 - \varepsilon < 1$ . Then (23) and (22) imply  $1 > \lambda(\tilde{z} - \delta\tilde{e} - \varepsilon\mathbf{1}) = \lambda(\tilde{z}) - \delta\lambda(\tilde{e}) - \varepsilon\lambda(\mathbf{1}) = 1 - \varepsilon - \delta\lambda(\tilde{e})$ . Hence,  $w_j^i = \lambda(\tilde{e}) > -\varepsilon/\delta$ , and the conclusion follows by letting  $\varepsilon$  go to 0.

Thirdly, we prove that  $\tilde{w}$  satisfies (30). It follows from Axiom C1 and (23) that

$$\forall c > 0, \lambda(\tilde{x}) < c \text{ for any } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x}) < c. \quad (31)$$

For any  $c \leq 0$ , we choose  $b > 0$  such that  $b + c > 0$ . Then by (31), we have  $\lambda(\tilde{x} + b\mathbf{1}) < c + b$  for any  $\tilde{x}$  such that  $\tilde{x} \in \mathcal{B}$  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

$$\forall c \leq 0, \lambda(\tilde{x}) < c \text{ for any } \tilde{x} \text{ such that } \tilde{x} \in \mathcal{B} \text{ and } \hat{\rho}(\tilde{x}) < c. \quad (32)$$

It follows from (31) and (32) that  $\hat{\rho}(\tilde{x}) \geq \lambda(\tilde{x})$  for any  $\tilde{x} \in \mathcal{B}$ , which in combination with  $\hat{\rho}(\tilde{z}) = 1 = \lambda(\tilde{z})$  completes the proof of (30).

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  $\frac{1}{\hat{\rho}(\tilde{z})}\tilde{z} \in \mathcal{B}^\circ$ , it follows from the result proved in Case 1 that there exists a weight  $\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^m) \in \mathbb{R}^n$  such that  $\tilde{w}$  satisfies (3), (4), and the linear functional  $\lambda(\tilde{x}) := \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i$  satisfies  $\hat{\rho}(\tilde{x}) \geq \lambda(\tilde{x})$  for  $\forall \tilde{x} \in \mathcal{B}$  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  $\hat{\rho}(\tilde{z}) = \lambda(\tilde{z})$ . Thus,  $\tilde{w}$  also satisfies (30).

Case 3:  $\hat{\rho}(\tilde{z}) \leq 0$ . Choose  $b > 0$  such that  $\hat{\rho}(\tilde{z} + b\mathbf{1}) > 0$ . Since  $\tilde{z} + b\mathbf{1} \in \mathcal{B}^\circ$ , it follows from the results proved in Case 1 and Case 2 that there exists a weight  $\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^m) \in \mathbb{R}^n$  such that  $\tilde{w}$  satisfies (3), (4), and the linear functional  $\lambda(\tilde{x}) := \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i$  satisfies  $\hat{\rho}(\tilde{x}) \geq \lambda(\tilde{x})$  for  $\forall \tilde{x} \in \mathcal{B}$ , and  $\hat{\rho}(\tilde{z} + b\mathbf{1}) = \lambda(\tilde{z} + b\mathbf{1})$ , or equivalently  $\hat{\rho}(\tilde{z}) = \lambda(\tilde{z})$ . Thus,  $\tilde{w}$  also satisfies (30).  $\square$

PROOF OF THEOREM 3.1. Firstly, we prove part (i). Suppose  $\hat{\rho}$  is defined by (5), then obviously  $\hat{\rho}$  satisfies Axiom C1 and C4. To check Axiom C2, suppose  $\tilde{x} \leq \tilde{y}$ . For each  $i = 1, \dots, m$ , let  $(p_{i,1}, \dots, p_{i,n_i})$  be the permutation of  $(1, \dots, n_i)$  such that  $(y_{(1)}^i, y_{(2)}^i, \dots, y_{(n_i)}^i) = (y_{p_{i,1}}^i, y_{p_{i,2}}^i, \dots, y_{p_{i,n_i}}^i)$ . Then for any  $1 \leq j \leq n_i$  and  $1 \leq i \leq m$ ,  $y_{(j)}^i = y_{p_{i,j}}^i = \max\{y_{p_{i,k}}^i; k = 1, \dots, j\} \geq \max\{x_{p_{i,k}}^i; k = 1, \dots, j\} \geq x_{(j)}^i$ , which implies that  $\hat{\rho}$  satisfies Axiom C2 because

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

To check Axiom C3, note that if  $\tilde{x}$  and  $\tilde{y}$  are scenario-wise comonotonic, then for each  $i = 1, \dots, m$ , there exists a permutation  $(p_{i,1}, \dots, p_{i,n_i})$  of  $(1, \dots, n_i)$  such that  $x_{p_{i,1}}^i \leq x_{p_{i,2}}^i \leq \dots \leq x_{p_{i,n_i}}^i$  and  $y_{p_{i,1}}^i \leq y_{p_{i,2}}^i \leq \dots \leq y_{p_{i,n_i}}^i$ . Hence, we have  $(\tilde{x}^i + \tilde{y}^i)_{(j)} = x_{p_{i,j}}^i + y_{p_{i,j}}^i = x_{(j)}^i + y_{(j)}^i$ ,  $j = 1, \dots, n_i; i = 1, \dots, m$ .



Therefore,

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

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

Secondly, we prove part (ii). Let  $\mathcal{B}$  be defined in (21). By Axiom C4, we only need to show that there exists a set of weights  $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^n$  such that each  $\tilde{w} \in \mathcal{W}$  satisfies condition (3) and (4), and  $\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i\}$  for  $\forall \tilde{x} \in \mathcal{B}$ .

By Lemma A.2, for any point  $\tilde{y} \in \mathcal{B}^\circ$ , there exists a weight  $\widetilde{w}(\tilde{y}) = (w(\tilde{y})_1^1, \dots, w(\tilde{y})_{n_1}^1; \dots; w(\tilde{y})_1^m, \dots, w(\tilde{y})_{n_m}^m) \in \mathbb{R}^n$  such that (3) and (4) hold, and that

$$\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i x_j^i \text{ for } \forall \tilde{x} \in \mathcal{B}, \text{ and } \hat{\rho}(\tilde{y}) = \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i y_j^i. \quad (33)$$

Define  $\mathcal{W}$  as the collection of such weights, i.e.,  $\mathcal{W} := \{\widetilde{w}(\tilde{y}) \mid \tilde{y} \in \mathcal{B}^\circ\}$ , then each  $\tilde{w} \in \mathcal{W}$  satisfies (3) and (4). From (33), for any fixed  $\tilde{x} \in \mathcal{B}^\circ$ , we have

$$\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i x_j^i \text{ for } \forall \tilde{y} \in \mathcal{B}^\circ, \text{ and } \hat{\rho}(\tilde{x}) = \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{x})_j^i x_j^i.$$

Therefore,

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{y} \in \mathcal{B}^\circ} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i x_j^i \right\} = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i \right\}, \quad \forall \tilde{x} \in \mathcal{B}^\circ. \quad (34)$$

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

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i \right\}, \quad \forall \tilde{x} \in \partial \mathcal{B}. \quad (35)$$

Let  $\tilde{b} = (b_1^1, \dots, b_{n_1}^1, \dots, b_1^m, \dots, b_{n_m}^m)$  be any boundary point of  $\mathcal{B}$ . Then there exists a sequence  $\{\tilde{b}(k)\}_{k=1}^\infty \subset \mathcal{B}^\circ$  such that  $\tilde{b}(k) \rightarrow \tilde{b}$  as  $k \rightarrow \infty$ . By the continuity of  $\hat{\rho}$  and (34), we have

$$\hat{\rho}(\tilde{b}) = \lim_{k \rightarrow \infty} \hat{\rho}(\tilde{b}(k)) = \lim_{k \rightarrow \infty} \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i b(k)_j^i \right\}. \quad (36)$$

If we can interchange sup and limit in (36), i.e. if

$$\lim_{k \rightarrow \infty} \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i b(k)_j^i \right\} = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \lim_{k \rightarrow \infty} \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i b(k)_j^i \right\} = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i b_j^i \right\}, \quad (37)$$

then (35) holds and the proof is completed. To show (37), note by Cauchy-Schwarz inequality

$$\begin{aligned} &\left| \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i b(k)_j^i - \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i b_j^i \right| \\ &\leq \left( \sum_{i=1}^m \sum_{j=1}^{n_i} (w_j^i)^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^m \sum_{j=1}^{n_i} (b(k)_j^i - b_j^i)^2 \right)^{\frac{1}{2}} \leq \left( \sum_{i=1}^m \sum_{j=1}^{n_i} (b(k)_j^i - b_j^i)^2 \right)^{\frac{1}{2}}, \quad \forall \tilde{w} \in \mathcal{W}, \end{aligned}$$

because  $w_j^i \geq 0$  and  $\sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i = 1, \forall \tilde{w} \in \mathcal{W}$ . Hence,  $\sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i b(k)_j^i \rightarrow \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i b_j^i$  uniformly for all  $\tilde{w} \in \mathcal{W}$  as  $k \rightarrow \infty$ . Therefore, (37) follows.  $\square$

**Appendix B. The Second Representation via Acceptance Sets.** A statistical acceptance set is a subset of  $\mathbb{R}^n$  that includes all the data considered acceptable by a regulator in terms of the risk measured from them. Given a statistical acceptance set  $\mathcal{A}$ , the risk statistic  $\hat{\rho}_{\mathcal{A}}$  associated with  $\mathcal{A}$  is defined to be

$$\hat{\rho}_{\mathcal{A}}(\tilde{x}) := \inf\{h \mid \tilde{x} - h\mathbf{1} \in \mathcal{A}\}, \quad \forall \tilde{x} \in \mathbb{R}^n. \quad (38)$$

$\hat{\rho}_{\mathcal{A}}(\tilde{x})$  is the minimum amount of cash that has to be added to the original position corresponding to  $\tilde{x}$  in order for the resulting position to be acceptable.

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

$$\mathcal{A}_{\hat{\rho}} := \{\tilde{x} \in \mathbb{R}^n \mid \hat{\rho}(\tilde{x}) \leq 0\}. \quad (39)$$

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

**Axiom D1.**  $\mathcal{A}$  contains  $\mathbb{R}^n$ , where  $\mathbb{R}^n := \{\tilde{x} \in \mathbb{R}^n \mid x_j^i \leq 0, j = 1, \dots, n_i; i = 1, \dots, m\}$ .

**Axiom D2.**  $\mathcal{A}$  does not intersect the set  $\mathbb{R}_{++}^n$ , where  $\mathbb{R}_{++}^n := \{\tilde{x} \in \mathbb{R}^n \mid x_j^i > 0, j = 1, \dots, n_i; i = 1, \dots, m\}$ .

**Axiom D3.** If  $\tilde{x}$  and  $\tilde{y}$  are scenario-wise 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.**  $\mathcal{A}$  is positively homogeneous, i.e., if  $\tilde{x} \in \mathcal{A}$ , then  $\lambda\tilde{x} \in \mathcal{A}$  for any  $\lambda \geq 0$ .

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

**Axiom D6.**  $\mathcal{A}$  is empirical-law-invariant, i.e., if  $\tilde{x} = (x_1^1, x_2^1, \dots, x_{n_1}^1, \dots, x_1^m, x_2^m, \dots, x_{n_m}^m) \in \mathcal{A}$ , then for any permutation  $(p_{i,1}, p_{i,2}, \dots, p_{i,n_i})$  of  $(1, 2, \dots, n_i)$ ,  $i = 1, \dots, m$ , it holds that  $(x_{p_{1,1}}^1, x_{p_{1,2}}^1, \dots, x_{p_{1,n_1}}^1, \dots, x_{p_{m,1}}^m, x_{p_{m,2}}^m, \dots, x_{p_{m,n_m}}^m) \in \mathcal{A}$ .

The following theorem shows that a natural risk statistic and a statistical acceptance set satisfying Axiom D1-D6 are mutually representable.

**THEOREM B.1** (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 (with  $s = 1$  in Axiom C1).

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

(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.** (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  $\hat{\rho}(0) \leq \hat{\rho}(\tilde{x} - \alpha\mathbf{1}) = \hat{\rho}(\tilde{x}) - \alpha s$ . So  $\hat{\rho}(\tilde{x}) \geq \alpha s > 0$  and hence  $\tilde{x} \notin \mathcal{A}_{\hat{\rho}}$ , i.e., D2 holds. (3) If  $\tilde{x}$  and  $\tilde{y}$  are scenario-wise 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 scenario-wise 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  $(p_{i,1}, p_{i,2}, \dots, p_{i,n_i})$  of  $(1, 2, \dots, n_i)$ ,  $i = 1, \dots, m$ , Axiom C4 implies  $\hat{\rho}((x_{p_{1,1}}^1, x_{p_{1,2}}^1, \dots, x_{p_{1,n_1}}^1, \dots, x_{p_{m,1}}^m, x_{p_{m,2}}^m, \dots, x_{p_{m,n_m}}^m)) = \hat{\rho}(\tilde{x}) \leq 0$ . So  $(x_{p_{1,1}}^1, x_{p_{1,2}}^1, \dots, x_{p_{1,n_1}}^1, \dots, x_{p_{m,1}}^m, x_{p_{m,2}}^m, \dots, x_{p_{m,n_m}}^m) \in \mathcal{A}_{\hat{\rho}}$ , i.e., D6 holds. (7) Suppose  $\{\tilde{x}(k)\}_{k=1}^{\infty} \subset \mathcal{A}_{\hat{\rho}}$ , and  $\tilde{x}(k) \rightarrow \tilde{x}$  as  $k \rightarrow \infty$ . Then  $\hat{\rho}(\tilde{x}(k)) \leq 0, \forall k$ . The continuity of  $\hat{\rho}$  (see the comment following the definition of Axiom C2) implies  $\hat{\rho}(\tilde{x}) = \lim_{k \rightarrow \infty} \hat{\rho}(\tilde{x}(k)) \leq 0$ . 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

$$\begin{aligned} \hat{\rho}_{\mathcal{A}}(\tilde{x} + b\mathbf{1}) &= \inf\{h \mid \tilde{x} + b\mathbf{1} - h\mathbf{1} \in \mathcal{A}\} = b + \inf\{h - b \mid \tilde{x} - (h - b)\mathbf{1} \in \mathcal{A}\} \\ &= b + \inf\{h \mid \tilde{x} - h\mathbf{1} \in \mathcal{A}\} = b + \hat{\rho}_{\mathcal{A}}(\tilde{x}). \end{aligned}$$

For  $\forall \tilde{x} \in \mathbb{R}^n, \forall a \geq 0$ , if  $a = 0$ , then  $\hat{\rho}_{\mathcal{A}}(a\tilde{x}) = \inf\{h \mid 0 - h\mathbf{1} \in \mathcal{A}\} = 0 = a\hat{\rho}_{\mathcal{A}}(\tilde{x})$ , where the second equality follows from Axiom D1 and D2. If  $a > 0$ , then

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

by Axiom D4. Therefore, C1 holds (with  $s = 1$ ). (2) Suppose  $\tilde{x} \leq \tilde{y}$ . For any  $h \in \mathbb{R}$ , if  $\tilde{y} - h\mathbf{1} \in \mathcal{A}$ , then Axiom D5 and  $\tilde{x} - h\mathbf{1} \leq \tilde{y} - h\mathbf{1}$  imply that  $\tilde{x} - h\mathbf{1} \in \mathcal{A}$ . Hence,  $\{h \mid \tilde{y} - h\mathbf{1} \in \mathcal{A}\} \subseteq \{h \mid \tilde{x} - h\mathbf{1} \in \mathcal{A}\}$ . 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 scenario-wise comonotonic. For any  $g$  and  $h$  such that  $\tilde{x} - g\mathbf{1} \in \mathcal{A}$  and  $\tilde{y} - h\mathbf{1} \in \mathcal{A}$ , since  $\tilde{x} - g\mathbf{1}$  and  $\tilde{y} - h\mathbf{1}$  are scenario-wise comonotonic, it follows from Axiom D3 that  $\frac{1}{2}(\tilde{x} - g\mathbf{1}) + \frac{1}{2}(\tilde{y} - h\mathbf{1}) \in \mathcal{A}$ . By Axiom D4, the previous formula implies  $\tilde{x} + \tilde{y} - (g + h)\mathbf{1} \in \mathcal{A}$ . Therefore,  $\hat{\rho}_{\mathcal{A}}(\tilde{x} + \tilde{y}) \leq g + h$ . Taking infimum of all  $g$  and  $h$  satisfying  $\tilde{x} - g\mathbf{1} \in \mathcal{A}, \tilde{y} - h\mathbf{1} \in \mathcal{A}$ , on both sides of the above inequality yields  $\hat{\rho}_{\mathcal{A}}(\tilde{x} + \tilde{y}) \leq \hat{\rho}_{\mathcal{A}}(\tilde{x}) + \hat{\rho}_{\mathcal{A}}(\tilde{y})$ . So C3 holds. (4) Fix any  $\tilde{x} \in \mathbb{R}^n$  and any permutation  $(p_{i,1}, p_{i,2}, \dots, p_{i,n_i})$  of  $(1, 2, \dots, n_i), i = 1, \dots, m$ . Then for any  $h \in \mathbb{R}$ , Axiom D6 implies that  $\tilde{x} - h\mathbf{1} \in \mathcal{A}$  if and only if  $(x_{p_{1,1}}^1, x_{p_{1,2}}^1, \dots, x_{p_{1,n_1}}^1, \dots, x_{p_{m,1}}^m, x_{p_{m,2}}^m, \dots, x_{p_{m,n_m}}^m) - h\mathbf{1} \in \mathcal{A}$ . Hence,  $\{h \mid \tilde{x} - h\mathbf{1} \in \mathcal{A}\} = \{h \mid (x_{p_{1,1}}^1, x_{p_{1,2}}^1, \dots, x_{p_{1,n_1}}^1, \dots, x_{p_{m,1}}^m, x_{p_{m,2}}^m, \dots, x_{p_{m,n_m}}^m) - h\mathbf{1} \in \mathcal{A}\}$ . Taking infimum, we obtain  $\hat{\rho}_{\mathcal{A}}(\tilde{x}) = \hat{\rho}_{\mathcal{A}}((x_{p_{1,1}}^1, x_{p_{1,2}}^1, \dots, x_{p_{1,n_1}}^1, \dots, x_{p_{m,1}}^m, x_{p_{m,2}}^m, \dots, x_{p_{m,n_m}}^m))$ , i.e., C4 holds.

(iii) For  $\forall \tilde{x} \in \mathbb{R}^n$ , we have  $\hat{\rho}_{\mathcal{A}_\rho}(\tilde{x}) = \inf\{h \mid \tilde{x} - h\mathbf{1} \in \mathcal{A}_\rho\} = \inf\{h \mid \hat{\rho}(\tilde{x} - h\mathbf{1}) \leq 0\} = \inf\{h \mid \hat{\rho}(\tilde{x}) \leq sh\} = \frac{1}{s}\hat{\rho}(\tilde{x})$ , where the third equality follows from Axiom C1.

(iv) For any  $\tilde{x} \in \mathcal{D}$ , we have  $\hat{\rho}_{\mathcal{D}}(\tilde{x}) \leq 0$ . Hence,  $\tilde{x} \in \mathcal{A}_{\hat{\rho}_{\mathcal{D}}}$ . Therefore,  $\mathcal{D} \subseteq \mathcal{A}_{\hat{\rho}_{\mathcal{D}}}$ . By the results (i) and (ii),  $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}}$  is closed. So  $\bar{\mathcal{D}} \subseteq \mathcal{A}_{\hat{\rho}_{\mathcal{D}}}$ . On the other hand, for any  $\tilde{x} \in \mathcal{A}_{\hat{\rho}_{\mathcal{D}}}$ , we have by definition that  $\hat{\rho}_{\mathcal{D}}(\tilde{x}) \leq 0$ , i.e.,  $\inf\{h \mid \tilde{x} - h\mathbf{1} \in \mathcal{D}\} \leq 0$ . If  $\inf\{h \mid \tilde{x} - h\mathbf{1} \in \mathcal{D}\} < 0$ , then there exists  $h < 0$  such that  $\tilde{x} - h\mathbf{1} \in \mathcal{D}$ . Then since  $\tilde{x} < \tilde{x} - h\mathbf{1}$ , by D5  $\tilde{x} \in \mathcal{D}$ . Otherwise,  $\inf\{h \mid \tilde{x} - h\mathbf{1} \in \mathcal{D}\} = 0$ . Then there exists  $h_k$  such that  $h_k \downarrow 0$  as  $k \rightarrow \infty$  and  $\tilde{x} - h_k\mathbf{1} \in \mathcal{D}$ . Hence,  $\tilde{x} \in \bar{\mathcal{D}}$ . In either case we obtain  $\tilde{x} \in \bar{\mathcal{D}}$ . Hence,  $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}} \subseteq \bar{\mathcal{D}}$ . Therefore, we conclude that  $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}} = \bar{\mathcal{D}}$ .  $\square$

**Appendix C. Proof of Theorem 3.3.** In this section, we assume without loss of generality that  $s = 1$  in Axiom C1. The proof for Theorem 3.3 follows the same line as that for Theorem 3.1. We first prove two lemmas that are similar to Lemma A.1 and A.2.

LEMMA C.1 *Let  $\mathcal{B}$  be defined in (21). 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  $\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^m) \in \mathbb{R}^n$  satisfying (10) such that the linear functional  $\lambda(\tilde{x}) := \sum_{j=1}^{n_1} w_j^1 x_j^1 + \sum_{j=1}^{n_2} w_j^2 x_j^2 + \dots + \sum_{j=1}^{n_m} w_j^m x_j^m$  satisfies*

$$\lambda(\tilde{z}) = 1, \tag{40}$$

$$\lambda(\tilde{x}) < 1 \text{ for any } \tilde{x} \text{ such that } \hat{\rho}(\tilde{x}) < 1. \tag{41}$$

PROOF. Let  $U = \{\tilde{x} \mid \hat{\rho}(\tilde{x}) < 1\}$ . Axiom C1 and E3 imply that  $U$  is convex, and, hence, 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  $\tilde{z} - \varepsilon\mathbf{1}$  converges to  $\tilde{z}$  as  $\varepsilon \downarrow 0$ ,  $\tilde{z}$  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  $\tilde{u} = (u_1^1, \dots, u_{n_1}^1, \dots, u_1^m, \dots, u_{n_m}^m) \in \mathbb{R}^n$  such that  $\mu(\tilde{x}) := \sum_{i=1}^m \sum_{j=1}^{n_i} u_j^i x_j^i$  satisfies  $\mu(\tilde{x}) \leq \mu(\tilde{z})$  for any  $\tilde{x} \in \bar{U}$ . In particular, we have

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

For each  $i = 1, \dots, m$ , let  $\phi_i : \{1, 2, \dots, n_i\} \rightarrow \{1, 2, \dots, n_i\}$  be a bijection such that  $u_{\phi_i(1)}^i \leq u_{\phi_i(2)}^i \leq \dots \leq u_{\phi_i(n_i)}^i$ , and  $\psi_i(\cdot)$  be the inverse of  $\phi_i(\cdot)$ . Define a new weight  $\tilde{w}$  and a new linear functional  $\lambda(\cdot)$  as follows:

$$w_j^i := u_{\phi_i(j)}^i, j = 1, \dots, n_i; i = 1, \dots, m, \tag{43}$$

$$\tilde{w} := (w_1^1, \dots, w_{n_1}^1, \dots, w_1^m, \dots, w_{n_m}^m), \tag{44}$$

$$\lambda(\tilde{x}) := \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i, \tag{45}$$

then by definition  $\tilde{w}$  satisfies condition (10). For any fixed  $\tilde{x} \in U$ , by Axiom C4,  $\hat{\rho}((x_{\psi_1(1)}^1, \dots, x_{\psi_1(n_1)}^1, \dots, x_{\psi_m(1)}^m, \dots, x_{\psi_m(n_m)}^m)) = \hat{\rho}(\tilde{x}) < 1$ , so  $(x_{\psi_1(1)}^1, \dots, x_{\psi_1(n_1)}^1, \dots, x_{\psi_m(1)}^m, \dots, x_{\psi_m(n_m)}^m) \in U$ . Then, we have

$$\begin{aligned} \lambda(\tilde{x}) &= \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i = \sum_{i=1}^m \sum_{j=1}^{n_i} u_{\phi_i(j)}^i x_j^i = \sum_{i=1}^m \sum_{j=1}^{n_i} u_{\phi_i(\psi_i(j))}^i x_{\psi_i(j)}^i = \sum_{i=1}^m \sum_{j=1}^{n_i} u_j^i x_{\psi_i(j)}^i \\ &= \mu((x_{\psi_1(1)}^1, \dots, x_{\psi_1(n_1)}^1, \dots, x_{\psi_m(1)}^m, \dots, x_{\psi_m(n_m)}^m)) \leq \mu(\tilde{z}) \text{ (by (42)).} \end{aligned} \quad (46)$$

Noting that  $z_1^i \leq z_2^i \leq \dots \leq z_{n_i}^i$ ,  $i = 1, 2, \dots, m$ , we obtain

$$\mu(\tilde{z}) = \sum_{i=1}^m \sum_{j=1}^{n_i} u_j^i z_j^i \leq \sum_{i=1}^m \sum_{j=1}^{n_i} u_{\phi_i(j)}^i z_j^i = \lambda(\tilde{z}). \quad (47)$$

By (46) and (47), we have

$$\lambda(\tilde{x}) \leq \lambda(\tilde{z}), \forall \tilde{x} \in U. \quad (48)$$

We shall show that the strict inequality holds in (48). Suppose, by contradiction, that there exists  $\tilde{r} \in U$  such that  $\lambda(\tilde{r}) = \lambda(\tilde{z})$ . With  $w_{\max} := \max\{w_1^1, \dots, w_{n_1}^1, \dots, w_1^m, \dots, w_{n_m}^m\}$ , we have  $\tilde{r} + \varepsilon \tilde{w} \leq \tilde{r} + \varepsilon w_{\max} \mathbf{1}$  for any  $\varepsilon > 0$ . Thus, Axiom C1 and C2 yield

$$\hat{\rho}(\tilde{r} + \varepsilon \tilde{w}) \leq \hat{\rho}(\tilde{r} + \varepsilon w_{\max} \mathbf{1}) = \hat{\rho}(\tilde{r}) + \varepsilon w_{\max}, \forall \varepsilon > 0. \quad (49)$$

Since  $\hat{\rho}(\tilde{r}) < 1$ , we have by (49) that for small enough  $\varepsilon > 0$ ,  $\hat{\rho}(\tilde{r} + \varepsilon \tilde{w}) < 1$ . Hence,  $\tilde{r} + \varepsilon \tilde{w} \in U$  and (48) implies  $\lambda(\tilde{r} + \varepsilon \tilde{w}) \leq \lambda(\tilde{z})$ . However,  $\lambda(\tilde{r} + \varepsilon \tilde{w}) = \lambda(\tilde{r}) + \varepsilon |\tilde{w}|^2 > \lambda(\tilde{r}) = \lambda(\tilde{z})$ , leading to a contradiction. In summary, we have shown that

$$\lambda(\tilde{x}) < \lambda(\tilde{z}), \forall \tilde{x} \in U. \quad (50)$$

Since  $\hat{\rho}(0) = 0$ , we have  $0 \in U$ . Letting  $\tilde{x} = 0$  in (50) yields  $\lambda(\tilde{z}) > 0$ , so we can re-scale  $\tilde{w}$  such that  $\lambda(\tilde{z}) = 1 = \hat{\rho}(\tilde{z})$ . Thus, (50) becomes  $\lambda(\tilde{x}) < 1$  for any  $\tilde{x}$  such that  $\hat{\rho}(\tilde{x}) < 1$ , from which (41) holds.  $\square$

**LEMMA C.2** *Let  $\mathcal{B}$  be defined in (21). For any fixed  $\tilde{z} \in \mathcal{B}$  and any  $\hat{\rho}$  satisfying Axiom C1, C2, E3 and C4, there exists a weight  $\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^m) \in \mathbb{R}^n$  satisfying (8), (9), and (10), such that*

$$\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i \text{ for any } \tilde{x} \in \mathbb{R}^n, \text{ and } \hat{\rho}(\tilde{z}) = \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i z_j^i. \quad (51)$$

**PROOF.** We will show this by considering two cases.

Case 1:  $\hat{\rho}(\tilde{z}) = 1$ . From Lemma C.1, there exists a weight  $\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^m) \in \mathbb{R}^n$  satisfying (10) such that the linear functional  $\lambda(\tilde{x}) := \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i$  satisfies (40) and (41).

Firstly, we prove that  $\tilde{w}$  satisfies (8), which is equivalent to  $\lambda(\mathbf{1}) = 1$ . To this end, first note that for any  $c < 1$ , Axiom C1 implies  $\hat{\rho}(c\mathbf{1}) = c < 1$ . Thus, (41) implies  $\lambda(c\mathbf{1}) < 1$ , and, by continuity of  $\lambda(\cdot)$ , we obtain that  $\lambda(\mathbf{1}) \leq 1$ . On the other hand, 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 (40) and (41) 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 (8). Secondly, we prove that  $\tilde{w}$  satisfies (9). For any fixed  $i$  and  $1 \leq j \leq n_i$ , let  $k = n_1 + n_2 + \dots + n_{i-1} + j$  and  $\tilde{e} = (0, \dots, 0, 1, 0, \dots, 0)$  be the  $k$ -th standard basis of  $\mathbb{R}^n$ . Then  $w_j^i = \lambda(\tilde{e})$ . For any  $\varepsilon > 0$ , Axiom C1 and C2 imply  $\hat{\rho}(\tilde{z} - \tilde{e} - \varepsilon\mathbf{1}) = \hat{\rho}(\tilde{z} - \tilde{e}) - \varepsilon \leq \hat{\rho}(\tilde{z}) - \varepsilon = 1 - \varepsilon < 1$ . Then (40) and (41) imply  $1 > \lambda(\tilde{z} - \tilde{e} - \varepsilon\mathbf{1}) = \lambda(\tilde{z}) - \lambda(\tilde{e}) - \varepsilon\lambda(\mathbf{1}) = 1 - \varepsilon - \lambda(\tilde{e})$ . Hence,  $w_j^i = \lambda(\tilde{e}) > -\varepsilon$ , and the conclusion follows by letting  $\varepsilon$  go to 0. Thirdly, we prove that  $\tilde{w}$  satisfies (51). It follows from Axiom C1 and (41) that

$$\forall c > 0, \lambda(\tilde{x}) < c \text{ for any } \tilde{x} \text{ such that } \hat{\rho}(\tilde{x}) < c. \quad (52)$$

For any  $c \leq 0$ , we choose  $b > 0$  such that  $b + c > 0$ . Then it follows from (52) that  $\lambda(\tilde{x} + b\mathbf{1}) < c + b$  for any  $\tilde{x}$  such that  $\hat{\rho}(\tilde{x} + b\mathbf{1}) < c + b$ . 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

$$\forall c \leq 0, \lambda(\tilde{x}) < c \text{ for any } \tilde{x} \text{ such that } \hat{\rho}(\tilde{x}) < c. \quad (53)$$

It follows from (52) and (53) that  $\hat{\rho}(\tilde{x}) \geq \lambda(\tilde{x})$  for any  $\tilde{x} \in \mathbb{R}^n$ , which in combination with  $\hat{\rho}(\tilde{z}) = 1 = \lambda(\tilde{z})$  completes the proof of (51).

Case 2:  $\hat{\rho}(\tilde{z}) \neq 1$ . The argument is the same as that in case 2 and case 3 of the proof for Lemma A.2.  $\square$

PROOF OF THEOREM 3.3. Without loss of generality, we assume  $s = 1$  in Axiom C1.

Firstly, we prove part (i). We only need to show that under condition (10), the risk statistic (11) satisfies subadditivity for any  $\tilde{x}$  and  $\tilde{y} \in \mathbb{R}^n$ . Let  $(p_{i,1}, p_{i,2}, \dots, p_{i,n_i})$  be the permutation of  $(1, \dots, n_i)$  such that  $(\tilde{x}^i + \tilde{y}^i)_{p_{i,1}} \leq (\tilde{x}^i + \tilde{y}^i)_{p_{i,2}} \leq \dots \leq (\tilde{x}^i + \tilde{y}^i)_{p_{i,n_i}}$ . Then for  $k = 1, \dots, n_i - 1$ , the partial sum up to  $k$  satisfies

$$\sum_{j=1}^k (\tilde{x}^i + \tilde{y}^i)_{(j)} = \sum_{j=1}^k (\tilde{x}^i + \tilde{y}^i)_{p_{i,j}} = \sum_{j=1}^k (x_{p_{i,j}}^i + y_{p_{i,j}}^i) \geq \sum_{j=1}^k (x_{(j)}^i + y_{(j)}^i). \quad (54)$$

In addition, we have for the total sum

$$\sum_{j=1}^{n_i} (\tilde{x}^i + \tilde{y}^i)_{(j)} = \sum_{j=1}^{n_i} (\tilde{x}^i + \tilde{y}^i)_j = \sum_{j=1}^{n_i} (x_j^i + y_j^i) = \sum_{j=1}^{n_i} (x_{(j)}^i + y_{(j)}^i). \quad (55)$$

Re-arranging the summation terms yields

$$\begin{aligned} \hat{\rho}(\tilde{x} + \tilde{y}) &= \hat{\rho}((\tilde{x}^1 + \tilde{y}^1, \tilde{x}^2 + \tilde{y}^2, \dots, \tilde{x}^m + \tilde{y}^m)) = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i (\tilde{x}^i + \tilde{y}^i)_{(j)} \right\} \\ &= \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \left[ \sum_{j=1}^{n_i-1} (w_j^i - w_{j+1}^i) \sum_{k=1}^j (\tilde{x}^i + \tilde{y}^i)_{(k)} + w_{n_i}^i \sum_{k=1}^{n_i} (\tilde{x}^i + \tilde{y}^i)_{(k)} \right] \right\}, \end{aligned}$$

which, along with (54), (55), and the fact  $w_j^i - w_{j+1}^i \leq 0$ , shows that

$$\begin{aligned} \hat{\rho}(\tilde{x} + \tilde{y}) &\leq \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \left[ \sum_{j=1}^{n_i-1} (w_j^i - w_{j+1}^i) \sum_{k=1}^j (x_{(k)}^i + y_{(k)}^i) + w_{n_i}^i \sum_{k=1}^{n_i} (x_{(k)}^i + y_{(k)}^i) \right] \right\} \\ &= \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_{(j)}^i + \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i y_{(j)}^i \right\} \\ &\leq \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_{(j)}^i \right\} + \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i y_{(j)}^i \right\} = \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y}). \end{aligned}$$

Secondly, we prove part (ii). Let  $\mathcal{B}$  be defined in (21). By Axiom C4, we only need to show that there exists a set of weights  $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^n$  such that each  $\tilde{w} \in \mathcal{W}$  satisfies (8), (9), and (10), and  $\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathcal{W}} \{\sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i\}$  for  $\forall \tilde{x} \in \mathcal{B}$ .

By Lemma C.2, for any  $\tilde{y} \in \mathcal{B}$ , there exists a weight  $\widetilde{w}(\tilde{y}) = (w(\tilde{y})_1^1, \dots, w(\tilde{y})_{n_1}^1; \dots; w(\tilde{y})_1^m, \dots, w(\tilde{y})_{n_m}^m)$  satisfying (8), (9), and (10), such that

$$\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i x_j^i \text{ for any } \tilde{x} \in \mathbb{R}^n, \text{ and } \hat{\rho}(\tilde{y}) = \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i y_j^i. \quad (56)$$

Define  $\mathcal{W}$  as the collection of such weights, i.e.,  $\mathcal{W} := \{\widetilde{w}(\tilde{y}) \mid \tilde{y} \in \mathcal{B}\}$ . Then each  $\tilde{w} \in \mathcal{W}$  satisfies (8), (9), and (10). From (56), for any fixed  $\tilde{x} \in \mathcal{B}$ , we have  $\hat{\rho}(\tilde{x}) \geq \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i x_j^i$  for  $\forall \tilde{y} \in \mathcal{B}$ , and  $\hat{\rho}(\tilde{x}) = \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{x})_j^i x_j^i$ . Therefore,

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{y} \in \mathcal{B}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i x_j^i \right\} = \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i \right\}, \quad \forall \tilde{x} \in \mathcal{B},$$

which completes the proof.  $\square$

**Appendix D. Proof of Theorem 3.4.** PROOF. We assume, without loss of generality, that  $s = 1$  in Axiom F4. The “if” part can be proved by using the same argument as that in the proof for part (i) of Theorem 3.1. To prove the “only if” part, we shall first prove

$$\hat{\rho}(c\tilde{x}) = c\hat{\rho}(\tilde{x}), \quad \forall c \geq 0, \forall \tilde{x} \geq 0. \quad (57)$$

By Axiom F3, we have  $\hat{\rho}(0) = \hat{\rho}(0) + \hat{\rho}(0)$ , so  $\hat{\rho}(0) = 0$ . Axiom F3 also implies  $\hat{\rho}(m\tilde{x}) = m\hat{\rho}(\tilde{x})$ ,  $\forall m \in \mathbb{N}$ ,  $\tilde{x} \in \mathbb{R}^n$ , and  $\hat{\rho}(\frac{k}{m}\tilde{x}) = \frac{1}{m}\hat{\rho}(k\tilde{x}) = \frac{k}{m}\hat{\rho}(\tilde{x})$ , for  $\forall m \in \mathbb{N}$ ,  $k \in \mathbb{N} \cup \{0\}$ ,  $\tilde{x} \in \mathbb{R}^n$ , or equivalently, for the set of nonnegative rational numbers  $\mathbb{Q}^+$ ,

$$\hat{\rho}(q\tilde{x}) = q\hat{\rho}(\tilde{x}), \quad \forall q \in \mathbb{Q}^+, \tilde{x} \in \mathbb{R}^n. \quad (58)$$

In general, for any  $c \geq 0$  there exist two sequences  $\{d_n\}_{n=1}^{\infty} \subset \mathbb{Q}^+$  and  $\{e_n\}_{n=1}^{\infty} \subset \mathbb{Q}^+$ , such that  $d_n \uparrow c$  and  $e_n \downarrow c$  as  $n \rightarrow \infty$ . Then for  $\forall \tilde{x} \geq 0$ ,  $\forall n$ , we have  $d_n\tilde{x} \leq c\tilde{x} \leq e_n\tilde{x}$ . It follows from Axiom F2 and (58) that  $d_n\hat{\rho}(\tilde{x}) = \hat{\rho}(d_n\tilde{x}) \leq \hat{\rho}(c\tilde{x}) \leq \hat{\rho}(e_n\tilde{x}) = e_n\hat{\rho}(\tilde{x})$ ,  $\forall n$ ,  $\forall \tilde{x} \geq 0$ . Letting  $n \rightarrow \infty$ , we obtain (57).

Now we are ready to prove the “only if” part. Let  $\tilde{e}_j := (0, \dots, 0, 1, 0, \dots, 0)$  be the  $j$ -th standard basis of  $\mathbb{R}^n$ , and  $\ell_1 := 0$ ,  $\ell_i := \sum_{j=1}^{i-1} n_j$ ,  $i = 2, \dots, m$ . By Axiom F1 and F3,

$$\begin{aligned} \hat{\rho}(\tilde{x}) &= \hat{\rho}((x_{(1)}^1, x_{(2)}^1, \dots, x_{(n_1)}^1, \dots, x_{(1)}^m, x_{(2)}^m, \dots, x_{(n_m)}^m)) \\ &= \hat{\rho}\left(\sum_{i=1}^m (0, 0, \dots, 0, x_{(1)}^i, x_{(2)}^i, \dots, x_{(n_i)}^i, 0, 0, \dots, 0)\right) \\ &= \sum_{i=1}^m \hat{\rho}((0, 0, \dots, 0, x_{(1)}^i, x_{(2)}^i, \dots, x_{(n_i)}^i, 0, 0, \dots, 0)). \end{aligned} \quad (59)$$

Further, by Axiom F3,

$$\begin{aligned} &\hat{\rho}((0, \dots, 0, x_{(1)}^i, x_{(2)}^i, \dots, x_{(n_i)}^i, 0, \dots, 0)) \\ &= \hat{\rho}((0, \dots, 0, 0, x_{(2)}^i - x_{(1)}^i, \dots, x_{(n_i)}^i - x_{(1)}^i, 0, \dots, 0)) + \hat{\rho}\left(x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) \\ &= \hat{\rho}((0, \dots, 0, 0, 0, x_{(3)}^i - x_{(2)}^i, \dots, x_{(n_i)}^i - x_{(2)}^i, 0, \dots, 0)) + \hat{\rho}\left((x_{(2)}^i - x_{(1)}^i) \sum_{j=\ell_i+2}^{\ell_i+n_i} \tilde{e}_j\right) \\ &\quad + \hat{\rho}\left(x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) \\ &= \dots \\ &= \hat{\rho}\left((x_{(n_i)}^i - x_{(n_i-1)}^i) \sum_{j=\ell_i+n_i}^{\ell_i+n_i} \tilde{e}_j\right) + \dots + \hat{\rho}\left((x_{(2)}^i - x_{(1)}^i) \sum_{j=\ell_i+2}^{\ell_i+n_i} \tilde{e}_j\right) + \hat{\rho}\left(x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) \\ &= (x_{(n_i)}^i - x_{(n_i-1)}^i)\hat{\rho}\left(\sum_{j=\ell_i+n_i}^{\ell_i+n_i} \tilde{e}_j\right) + \dots + (x_{(2)}^i - x_{(1)}^i)\hat{\rho}\left(\sum_{j=\ell_i+2}^{\ell_i+n_i} \tilde{e}_j\right) + \hat{\rho}\left(x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right), \end{aligned} \quad (60)$$

where the last equality follows from (57). If  $x_{(1)}^i \geq 0$ , then by (57) we have

$$\hat{\rho}\left(x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) = x_{(1)}^i \hat{\rho}\left(\sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right). \quad (61)$$

If  $x_{(1)}^i < 0$ , then since  $x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j$  and  $-x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j$  are scenario-wise comonotonic, we have by Axiom F3 that  $\hat{\rho}\left(x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) + \hat{\rho}\left(-x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) = \hat{\rho}(0) = 0$ , which implies

$$\hat{\rho}\left(x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) = -\hat{\rho}\left(-x_{(1)}^i \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) = x_{(1)}^i \hat{\rho}\left(\sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right), \quad (62)$$

where the last equality follows from (57). Then by (59), (60), (61), and (62), we obtain  $\hat{\rho}(\tilde{x}) = \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_{(j)}^i$ , where  $w_j^i := \hat{\rho}(\sum_{k=\ell_i+j}^{\ell_i+n_i} \tilde{e}_k) - \hat{\rho}(\sum_{k=\ell_i+j+1}^{\ell_i+n_i} \tilde{e}_k)$ . Since by Axiom F2  $w_j^i \geq 0$  and  $\sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i = \hat{\rho}(\mathbf{1}) = 1$ , the proof is completed.  $\square$

**Appendix E. Proof of Theorem 4.1 and Theorem 4.2.** PROOF OF THEOREM 4.1. Let  $\tilde{x}^i = (x_1^i, \dots, x_{n_i}^i) \in \mathbb{R}^{n_i}$  be the data subset that is used to calculate  $\text{VaR}_{t-i}$ ,  $i = 1, \dots, 60$ , and  $\tilde{x}^{i+60} = (x_1^{i+60}, \dots, x_{n_{i+60}}^{i+60}) \in \mathbb{R}^{n_{i+60}}$  be the data subset used to calculate  $\text{sVaR}_{t-i}$ ,  $i = 1, \dots, 60$ . In addition, define the 121-th scenario  $\tilde{x}^{121} := 0 \in \mathbb{R}$  and  $n_{121} := 1$ . Let  $n := \sum_{i=1}^{121} n_i$ . We will show that (14) and (15) are natural risk statistics defined on  $\mathbb{R}^n$ . Define  $\tilde{w} = (\tilde{w}^1, \dots, \tilde{w}^{121}) = (w_1^1, \dots, w_{n_1}^1, \dots, w_1^{121}, \dots, w_{n_{121}}^{121}) \in \mathbb{R}^n$  such that  $w_j^i := 1_{\{j=\lceil 0.99n_i \rceil\}}$ ,  $1 \leq j \leq n_i$ ,  $i = 1, \dots, 121$ . Then we have

$$\text{VaR}_{t-i} = \sum_{j=1}^{n_i} w_j^i x_{(j)}^i, \quad \text{sVaR}_{t-i} = \sum_{j=1}^{n_{i+60}} w_j^{i+60} x_{(j)}^{i+60}; \quad i = 1, \dots, 60. \quad (63)$$

By (63), the Basel II risk measure (14) is equal to

$$k \cdot \max \left\{ \sum_{i=1}^{121} \sum_{j=1}^{n_i} u_j^i x_{(j)}^i, \sum_{i=1}^{121} \sum_{j=1}^{n_i} v_j^i x_{(j)}^i \right\}, \quad (64)$$

where the two weights  $\tilde{u} = (\tilde{u}^1, \dots, \tilde{u}^{121})$  and  $\tilde{v} = (\tilde{v}^1, \dots, \tilde{v}^{121})$  are defined by

$$\begin{aligned} \tilde{u}^1 &:= \frac{1}{k} \tilde{w}^1; \quad \tilde{u}^i := 0, i = 2, \dots, 120; \quad \tilde{u}^{121} := \frac{k-1}{k} \tilde{w}^{121}, \\ \tilde{v}^i &:= \frac{1}{60} \tilde{w}^i, i = 1, \dots, 60; \quad \tilde{v}^i := 0, i = 61, \dots, 121. \end{aligned}$$

Hence, by Theorem 3.1, (14) is a natural risk statistic that corresponds to  $s = k$  in Axiom C1. Again, by (63), the Basel III risk measure (15) is equal to

$$k \cdot \max \left\{ \sum_{i=1}^{121} \sum_{j=1}^{n_i} u_j^i x_{(j)}^i, \sum_{i=1}^{121} \sum_{j=1}^{n_i} v_j^i x_{(j)}^i \right\} + \ell \cdot \max \left\{ \sum_{i=1}^{121} \sum_{j=1}^{n_i} g_j^i x_{(j)}^i, \sum_{i=1}^{121} \sum_{j=1}^{n_i} h_j^i x_{(j)}^i \right\}, \quad (65)$$

where the two weights  $\tilde{g} = (\tilde{g}^1, \dots, \tilde{g}^{121})$  and  $\tilde{h} = (\tilde{h}^1, \dots, \tilde{h}^{121})$  are defined by

$$\begin{aligned} \tilde{g}^i &:= 0, \forall i \neq 61 \text{ and } i \neq 121; \quad \tilde{g}^{61} := \frac{1}{\ell} \tilde{w}^{61}; \quad \tilde{g}^{121} = \frac{\ell-1}{\ell} \tilde{w}^{121}, \\ \tilde{h}^i &:= 0, i = 1, \dots, 60; \quad \tilde{h}^i := \frac{1}{60} \tilde{w}^i, i = 61, \dots, 120; \quad \tilde{h}^{121} = 0. \end{aligned}$$

It is straightforward to verify that (65) satisfies Axiom C1-C4, with  $s = k + \ell$  in Axiom C1. Hence, (15) is also a natural risk statistic.  $\square$

PROOF OF THEOREM 4.2. The IRC risk measure (16) corresponds to a natural risk statistic with  $s = 1$ , which can be shown by following the same argument as that for proving Theorem 4.1.  $\square$

**Appendix F. Analysis of the Robustness of TCM and TCE.** The following tools in robust statistics show that TCM is more robust than TCE.

(i) The *influence function* is an important tool for assessing the robustness of statistics. Let  $F$  be the distribution function of  $X$ ,  $\tilde{x} = (x_1, \dots, x_n)$  be a sample of  $X$ , and  $F_n(\cdot)$  be the empirical distribution function. Let  $\mathcal{M}$  be the space of distribution functions on  $\mathbb{R}$ . Consider estimating  $T(F)$  from  $\tilde{x}$  for some statistical functional  $T(\cdot) : \mathcal{M} \rightarrow \mathbb{R}$ . TCM and TCE are both such functionals, since

$$\text{TCM}_\alpha(F) = F^{-1}\left(\frac{1+\alpha}{2}\right), \quad \text{TCE}_\alpha(F) = \frac{1}{1-\alpha} \int_\alpha^1 F^{-1}(s) ds. \quad (66)$$

A natural estimator for  $T(F)$  is  $T(F_n)$ , e.g.,

$$\text{TCM}_\alpha(F_n) = x_{(\lceil n(1+\alpha)/2 \rceil)}, \quad (67)$$

$$\text{TCE}_\alpha(F_n) = \frac{1}{1-\alpha} \left[ \left(\frac{k}{n} - \alpha\right) x_{(k)} + \frac{1}{n} \sum_{j=k}^{n-1} x_{(j+1)} \right], \quad k = \lceil n\alpha \rceil. \quad (68)$$

The robustness of the statistic  $T(F_n)$  can be asymptotically characterized by its influence function  $IF(y, T, F) := \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} [T((1-\varepsilon)F + \varepsilon\delta_y) - T(F)]$ ,  $y \in \mathbb{R}$ , where  $\delta_y$  is the point mass 1 at  $y$  which represents a contamination point to the distribution  $F$ . If the influence function is bounded, i.e.,  $\sup_{y \in \mathbb{R}} |IF(y, T, F)| < \infty$ , then  $T(F_n)$  is robust; otherwise,  $T(F_n)$  is not robust and outliers in the data may cause large changes to  $T(F_n)$  (Huber and Ronchetti [25]). The following proposition shows that TCE has an unbounded influence function but TCM has a bounded one.

PROPOSITION F.1 *If the random loss  $X$  has a positive probability density function  $f(\cdot)$ , then the influence function of  $TCM_\alpha$  is given by*

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

If  $f(\cdot)$  is continuous at  $VaR_\alpha(X)$ , then the influence function of  $TCE_\alpha$  is given by

$$IF(y, TCE_\alpha, X) = \begin{cases} VaR_\alpha(X) - TCE_\alpha(X), & \text{if } y \leq VaR_\alpha(X), \\ \frac{y}{1-\alpha} - TCE_\alpha(X) - \frac{\alpha}{1-\alpha} VaR_\alpha(X), & \text{if } y > VaR_\alpha(X). \end{cases} \quad (69)$$

PROOF. The result for TCM can be found at Section 5.3.1 of Shao [41]. To show (69), noting that by equation (3.2.4) in Staudte and Sheather [44], the influence function of the  $(1 - \alpha)$ -trimmed mean  $T_{1-\alpha}(X) := E[X|X < VaR_\alpha(X)]$  is

$$IF(y, T_{1-\alpha}, X) = \begin{cases} \frac{y - (1-\alpha)VaR_\alpha(X)}{\alpha} - E[X|X < VaR_\alpha(X)], & \text{if } y \leq VaR_\alpha(X), \\ VaR_\alpha(X) - E[X|X < VaR_\alpha(X)], & \text{if } y > VaR_\alpha(X). \end{cases} \quad (70)$$

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

$$IF(y, E[X], X) = y - E[X]. \quad (71)$$

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

$$IF(y, E[X], X) = \alpha IF(y, T_{1-\alpha}, X) + (1 - \alpha)IF(y, TCE_\alpha, X). \quad (72)$$

Now (69) follows from equations (70), (71), and (72).  $\square$

(ii) The *asymptotic breakdown point* is, roughly, the smallest fraction of bad observations that may cause an estimator to take on arbitrarily large aberrant values; see Section 1.4 of Huber and Ronchetti [25] for the mathematical definition. Hence, a high breakdown point is clearly desirable. From (66), both TCM and TCE are L-statistics, and then by Theorem 3.7 in Huber and Ronchetti [25], the asymptotic breakdown point for  $TCM_\alpha$  is  $1 - \alpha$  and that for  $TCE_\alpha$  is 0, which clearly shows the robustness of TCM.

(iii) Hampel's theorem (see Huber and Ronchetti [25, Section 2.6]) shows that for a statistical functional  $T(\cdot) : \mathcal{M} \rightarrow \mathbb{R}$ , the sequence of statistics  $T(F_n)$  is robust at  $F_0$  if and only if  $T$  is continuous at  $F_0$ . Again by Theorem 3.7 in Huber and Ronchetti [25],  $TCM_\alpha$  is continuous at  $F_0$  provided  $F_0^{-1}$  is continuous at  $(1 + \alpha)/2$ , but  $TCE_\alpha$  is discontinuous.

(iv) The *finite sample breakdown point* (see Huber and Ronchetti [25, Chap. 11]) of  $TCM_\alpha(F_n)$  is  $(n - \lceil n(1 + \alpha)/2 \rceil + 1)/(2n - \lceil n(1 + \alpha)/2 \rceil + 1) \approx (1 - \alpha)/(3 - \alpha)$ , but that of  $TCE_\alpha(F_n)$  is  $1/(n + 1)$ , which means one additional corrupted sample can cause arbitrarily large bias to  $TCE_\alpha$ .

**Acknowledgments.** We thank many people who offered insights into this work, including John Birge, Mark Broadie, Louis Eeckhoudt, Marco Frittelli, Paul Glasserman, Michael B. Gordy, and Jeremy Staum. We particularly thank two anonymous referees for their constructive comments that help to improve the paper. We have also benefited from the comments of seminar participants at Columbia University, Cornell University, Fields Institute, Georgia State University, Hong Kong University of Science and Technology, Stanford University, the University of Alberta, the University of Michigan, and of conference participants at INFORMS annual meetings. Steven Kou is supported in part by the U.S. National Science Foundation. Xianhua Peng is partially supported by a grant from School-Based-Initiatives of HKUST (Project No. SBI11SC03). The preliminary versions of the paper were entitled “What is a good (external) risk measure: Bridging the gaps between data (robustness), coherent risk measures (subadditivity), and insurance risk measures”.

## References

- [1] Acerbi, C., D. Tasche. 2002. On the coherence of expected shortfall. *J. Bank. Finance* **26** 1487–1503.
- [2] Adrian, T., M. K. Brunnermeier. 2008. CoVaR. *Federal Reserve Bank of New York Staff Reports No. 348*.
- [3] Ahmed, S., D. Filipović, G. Svindland. 2008. A note on natural risk statistics. *Oper. Res. Lett.* **36** 662–664.



- [4] Artzner, P., F. Delbaen, J.-M. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Finance* **9** 203–228.
- [5] Basel Committee on Banking Supervision. 2006. *International Convergence of Capital Measurement and Capital Standards: A Revised Framework (Comprehensive Version)*.
- [6] Basel Committee on Banking Supervision. July 2009. *Guidelines for Computing Capital for Incremental Risk in the Trading Book*.
- [7] Basel Committee on Banking Supervision. February 2011. *Revisions to the Basel II Market Risk Framework*.
- [8] Daniélsson, J., B. N. Jorgensen, G. Samorodnitsky, M. Sarma, C. G. de Vries. 2005. Subadditivity re-examined: the case for Value-at-Risk. Working paper, London School of Economics.
- [9] Delbaen, F. 2002. Coherent risk measures on general probability spaces. In: Sandmann, K., Schönbucher, P. J. (Eds.), *Advances in Finance and Stochastics—Essays in Honour of Dieter Sondermann*. Springer, New York.
- [10] Denneberg, D. 1994. *Non-Additive Measure and Integral*. Kluwer Academic Publishers, Boston.
- [11] Dhaene, J., M. J. Goovaerts, R. Kaas. 2003. Economic capital allocation derived from risk measures. *N. Am. Actuar. J.* **7** 44–59.
- [12] Eling, M., L. Tibiletti. 2010. Internal vs. external risk measures: how capital requirements differ in practice. *Oper. Res. Lett.* **38** 482–488.
- [13] Föllmer, H., A. Schied. 2002. Convex measures of risk and trading constraints. *Finance Stoch.* **6** 429–447.
- [14] Frittelli, M., E. R. Gianin. 2002. Putting order in risk measures. *J. Bank. Finance* **26** 1473–1486.
- [15] Garcia, R., É. Renault, G. Tsafack. 2007. Proper conditioning for coherent VaR in portfolio management. *Management Sci.* **53** 483–494.
- [16] Gilboa, I., D. Schmeidler. 1989. Maxmin expected utility with non-unique prior. *J. Math. Econom.* **18** 141–153.
- [17] Glasserman, P. 2009. Risk horizon and rebalancing horizon in portfolio risk measurement. *Math. Finance*, forthcoming.
- [18] Gordy, M. B. 2003. A risk-factor model foundation for ratings-based bank capital rules. *J. Financial Intermed.* **12** 199–232.
- [19] Gordy, M. B., B. Howells. 2006. Procyclicality in Basel II: can we treat the disease without killing the patient? *J. Financial Intermed.* **15** 395–417.
- [20] Hansen, L. P., T. J. Sargent. 2007. *Robustness*. Princeton University Press.
- [21] Hart, H. L. A. 1994. *The Concept of Law*. 2nd ed. Clarendon Press, Oxford.
- [22] Heyde, C. C., S. G. Kou. 2004. On the controversy over tailweight of distributions. *Oper. Res. Lett.* **32** 399–408.
- [23] Hong, C.-S., L. F. Herk. 1996. Increasing risk aversion and diversification preference. *J. Econom. Theory* **70** 180–200.
- [24] Huber, P. J. 1981. *Robust Statistics*. Wiley, New York.
- [25] Huber, P. J., E. M. Ronchetti. 2009. *Robust Statistics*. 2nd ed. Wiley, New York.
- [26] Ibragimov, R., 2004a. On the robustness of economic models to heavy-tailedness assumptions. Mimeo, Yale University.
- [27] Ibragimov, R., 2004b. Portfolio diversification and value at risk under thick-tailedness. Harvard Institute of Economic Research Discussion Paper #2086.
- [28] Ibragimov, R., J. Walden. 2007. The limits of diversification when losses may be large. *J. Bank. Finance* **31** 2551–2569.
- [29] Jaschke, S., U. Küchler. 2001. Coherent risk measures and good deal bounds. *Finance Stoch.* **5** 181–200.
- [30] Jorion, P. 2007. *Value at Risk: The New Benchmark for Managing Financial Risk*. 3rd ed. McGraw-Hill, Boston.
- [31] Keppo, J., L. Kofman, X. Meng. 2010. Unintended consequences of the market risk requirement in banking regulation. Working paper, University of Michigan.
- [32] Kijima, M. 1997. The generalized harmonic mean and a portfolio problem with dependent assets. *Theory and Decision* **43** 71–87.
- [33] Kusuoka, S. 2001. On law invariant coherent risk measures. *Adv. Math. Econom.* **3** 83–95.
- [34] Litterman, R. 1996. Hot spots<sup>TM</sup> and hedges. *J. Portfolio Management* **22** 52–75.
- [35] Maccheroni, F., M. Marinacci, A. Rustichini. 2006. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica* **74** 1447–1498.
- [36] McMin, R. D. 1984. A general diversification theorem: a note. *J. Finance* **39** 541–550.
- [37] McNeil, A., R. Frey, P. Embrechts. 2005. *Quantitative Risk Management*. Princeton University Press.
- [38] Rockafellar, R. T., S. Uryasev. 2002. Conditional Value-at-Risk for general loss distributions. *J. Bank. Finance* **26** 1443–1471.
- [39] Samuelson, P. A. 1967. General proof that diversification pays. *J. Financial Quant. Anal.* **2** 1–13.

- [40] Schmeidler, D. 1989. Subjective probability and expected utility without additivity. *Econometrica* **57** 571–587.
- [41] Shao, J. 2003. *Mathematical Statistics*. 2nd ed. Springer-Verlag.
- [42] Song, Y., J.-A. Yan. 2006. The representations of two types of functionals on  $L^\infty(\Omega, \mathcal{F})$  and  $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ . *Sci. China Ser. A: Math.* **49** 1376–1382.
- [43] Song, Y., J.-A. Yan. 2006. Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. Working paper, Chinese Academy of Sciences.
- [44] Staudte, R. G., S. J. Sheather. 1990. *Robust Estimation and Testing*. Wiley, New York.
- [45] Staum, J. 2004. Fundamental theorems of asset pricing for good deal bounds. *Math. Finance* **14** 141–161.
- [46] Tasche, D. 1999. Risk contributions and performance measurement. Preprint. Technische Universität München.
- [47] Transportation Research Board of the National Academies. 2003. *Design Speed, Operating Speed, and Posted Speed Practices*.
- [48] Wang, S. S., V. R. Young, H. H. Panjer. 1997. Axiomatic characterization of insurance prices. *Insurance Math. Econom.* **21** 173–183.
- [49] Yaari, M. E. 1987. The dual theory of choice under risk. *Econometrica* **55** 95–115.