Comments on the Consultative Document "Fundamental Review of the Trading Book: A Revised Market Risk Framework" Released by Bank for International Settlement in October, 2013

Steven Kou and Xianhua Peng

National University of Singapore, Columbia University, and Hong Kong University of Science and Technology

matsteve@nus.edu.sg, maxhpeng@ust.hk

Elementary statistics teaches us that both mean and median measure the average size of a random quantity, but they have different properties. In particular, if we want to obtain a robust measurement, then median is a better choice than mean. Now what does this have to do with trading book capital requirements?

It is proposed in the consultative document that one of the major changes to the trading book capital rule is to move from value-at-risk (VaR) to expected shortfall (ES) mainly because of "the inability of the measure [VaR] to capture the tail risk of the loss distribution."

We fully agree that it is necessary to capture the tail risk beyond the loss level specified by VaR. However, how to achieve this is debatable. More precisely, should we use ES, defined as the mean of the size of the loss beyond VaR (as suggested in the document), or median shortfall (MS), defined as the median of the size of the loss beyond VaR (Kou, Peng, Heyde, 2013)? This is related to the question about choosing between mean and median.

For example, if we want to capture the tail risk, e.g., the size of the loss beyond VaR at 99% level, we can either use ES at 99% level, which is the mean of the size of the loss beyond VaR at 99% level, or, alternatively, median shortfall at 99% level, which is the median of the size of the loss beyond VaR at 99%. Hence, just like ES, median shortfall also measures the riskiness of random losses by taking into account both the size and likelihood of losses. However, median shortfall has several advantages over the expected shortfall, in face of statistical model uncertainty.

Model Uncertainty

In the internal models-based approach for determining trading book capital requirements, regulators impose the risk measure and allow institutions to use their own internal risk models and private data in the calculation. There can be several statistically indistinguishable models for the same instrument or portfolio due to limited availability of data. In particular, the heaviness of tail distributions cannot be identified in many cases. For example, Heyde and Kou (2004) show that it is very difficult to distinguish between exponential-type and power-type tails with 5,000 observations (about 20 years of daily observations) because the quantiles of the two types of

distributions may overlap. Therefore, the tail behavior may be a subjective issue depending on people's modeling preferences.

The First Advantage of Median Shortfall: Elicitability

Median shortfall satisfies a basic statistical property called elicitability (i.e. there exists an objective function such that minimizing the expected objective function yields the risk measure; see Gneiting, 2011), but ES does not. If a risk measure is not elicitable, then it is hard to justify the use of a forecasting procedure for the risk measure.

More precisely, in face of model uncertainty, several forecasting procedures based on different models for the underlying risk can be used to forecast the risk measure. It is hence desirable to be able to evaluate which procedure gives a better forecast. The elicitability of a risk measure means that the risk measure can be obtained by minimizing the expectation of a forecasting objective function; hence, the forecasting objective function can be used for evaluating different forecasting procedures. On the other hand, if one cannot find such a forecasting objective function, then one cannot tell which one of competing point forecasts for the risk measurement performs the best by comparing their forecasting error, no matter what objective function is used.

In fact, the non-elicitability of ES "may challenge the use of ES as a predictive measure of risk, and may provide a partial explanation for the lack of literature on the evaluation of ES forecasts" (Gneiting, 2011).

On the contrary, median shortfall is elicitable (Gneiting, 2011, Kou and Peng, 2014).

The Second Advantage of Median Shortfall: Robustness

Median shortfall has the desirable property of distributional robustness with respective to model misspecification in the sense of Hampel (1971), which means that a small deviation of the model only results in a small change in the risk measurement; but ES does not (Kou, Peng, Heyde, 2013; Kou and Peng, 2014). This means that median shortfall leads to "more stable model output and often less sensitivity to extreme outlier observations," a desirable property mentioned on p. 18 of the current consultative document.

To further compare the robustness of MS with ES, Kou and Peng (2014) carry out a simple empirical study on the measurement of the tail risk of S&P 500 daily return. They consider two IGARCH(1, 1) models similar to the model of RiskMetrics, one with the noise having the standard normal distribution and the other t-distribution with the degree of freedom unknown. After fitting the two models to the historical data of daily returns of S&P 500 Index during 1/2/1980–11/26/2012 and then forecasting the one-day median shortfall and ES of a portfolio of S&P500 stocks, it is found that the change of ES under the two models is much larger than that of median shortfall, indicating that ES is more sensitive to model misspecification than MS.

Regulatory risk measures should demonstrate robustness with respect to model misspecification (Kou, Peng, Heyde, 2013). From a regulator's viewpoint, a regulatory 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 regulatory 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 regulatory risk measure is not robust, institutions can take regulatory arbitrage by choosing a model that significantly reduces the capital requirements.

The requirement of robustness for regulatory risk measures is not anything new; in general, robustness is essential for law enforcement, as is implied by legal realism, one of the basic concepts of law; see Hart (1994). Legal realism is the viewpoint that a law is only a guideline for judges and enforcement officers (Hart, pp. 204–205) and is only intended to be the average of what judges and officers will decide. Hence, a law should be established in a robust way so that different judges will reach similar conclusions when they implement the law. In particular, the risk measures imposed in banking regulation should also be robust with respect to underlying models and data.

It is also worth noting that it is not desirable for a risk measure to be too sensitive to the tail risk. For example, consider the random loss that could occur to a person who walks on the street. There is a very small but positive probability that the person could be hit by a car and lose his life; in that unfortunate case, the loss may be infinite. Hence, the ES of the random loss may be equal to infinity, suggesting that the person should never walk on the street, which is apparently not reasonable. In contrast, the MS of the random loss is a finite number.

The Third Advantage of Median Shortfall: Easy Implementation

Kou and Peng (2014) show that, for any loss distribution, median shortfall at a given confidence level is simply equal to VaR at a higher confidence level. For example, median shortfall at 99% level is simply equal to VaR at 99.5% level.

Furthermore, the backtesting for median shortfall can be easily done using the existing methods for backtesting VaR (see, e.g., Jorion 2007; Gaglianone, et al. 2011), while it is difficult to do backtesting for ES. In fact, the current document suggests to do backtesting by "comparing 1-day static value-at-risk measure at both the 97.5th percentile and the 99th percentile to actual P&L outcomes", although it suggests to replace VaR by ES in measuring risk.

Conclusion

It is better to use median shortfall than ES. In fact, Kou and Peng (2014) prove that median shortfall is the only tail risk measure that satisfies a set of axioms based on the Choquet expected utility theory (Schmeidler, 1989) and has the statistical property of elicitability. Furthermore, median shortfall is robust with respect to model misspecification. Expected shortfall is neither elicitable nor robust; and it is difficult to implement ES and to do backtesting for ES.

In the last decade financial institutions around the globe have spent considerable effort to develop capacities to compute VaR. Implementation of median shortfall is as easy as VaR, as median

shortfall can be computed as VaR at a higher level. Shifting from VaR to ES, as proposed in the current document, not only lacks sound justification, but may also lead to huge implementation problems in financial institutions.

In short, median shortfall is a better alternative than expected shortfall as a risk measure for setting capital requirements in the Basel Accord.

References

Gaglianone, W. P., L. R. Lima, O. Linton, D. R. Smith (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics,* Vol. 29, 150–160.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, Vol. 106, 746–762.

Hart, H. (1994). *The Concept of Law*. 2nd ed., Clarenton Press, Oxford.

Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annuals of Mathematical Statistics,* Vol. 42, 1887–1896.

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

Jorion, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk*, 3 ed., McGraw-Hill, Boston.

Kou, S., X. Peng, and C. C. Heyde (2013). External risk measures and Basel accords. *Mathematics of Operations Research*, Vol. 38, 393–417.

Kou, S. and X. Peng (2014). On the measurement of economic tail risk. Preprint. National University of Singapore, Columbia University, and Hong Kong University of Science and Technology.

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

On the Measurement of Economic Tail Risk*

Steven Kou† Xianhua Peng‡

This version January 2014

Abstract

This paper attempts to provide a decision theoretical foundation for the measurement of economic tail risk, which is not only closely related to utility theory but also relevant to statistical model uncertainty. The main result is that the only tail risk measure that satisfies a set of economic axioms proposed by Schmeidler (1989, Econometrica) and the statistical property of elicitability (i.e. there exists an objective function such that minimizing the expected objective function yields the risk measure; see Gneiting 2011, J. Amer. Stat. Assoc.) is median shortfall, which is the median of tail loss distribution. Elicitability is important for backtesting. Median shortfall has the desirable property of distributional robustness with respective to model misspecification. We also extend the result to address model uncertainty by incorporating multiple scenarios. As an application, we argue that median shortfall is a better alternative than expected shortfall for setting capital requirements in Basel Accords.

Keywords: comonotonic independence, model uncertainty, robustness, elicitability, backtest, Value-at-Risk

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

*We thank Ruodu Wang for helpful comments. We are also grateful to the seminar and conference participants at National University of Singapore, Peking University, University of Waterloo, Risk Minds Asia Conference 2013, Quantitative Methods in Finance Conference 2013, INFORMS Annual Meeting 2013, and Risk and Regulation Workshop 2014 for their helpful comments and discussion. This research is supported by the University Grants Committee of HKSAR of China, and the Department of Mathematics of HKUST. This research was partially completed while Xianhua Peng was visiting the Institute for Mathematical Sciences and the Center for Quantitative Finance, National University of Singapore, in 2013.

†Department of Mathematics, National University of Singapore and Department of Industrial Engineering and Operations Research, Columbia University (on leave).

‡Corresponding author. Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Email: maxhpeng@ust.hk.

1 Introduction

This paper attempts to provide a decision theoretical foundation for the measurement of economic tail risk. Two important applications are setting insurance premiums and capital requirements for financial institutions. For example, a widely used class of risk measures for setting insurance risk premiums is proposed by Wang, Young and Panjer (1997) based on a set of axioms. In terms of capital requirements, Gordy (2003) 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% Value-at-Risk (VaR). VaR is a widely used approach for the measurement of tail risk; the estimation and backtesting of VaR has been well studies in the literature, see, e.g., Duffie and Pan (1997, 2001); Jorion (2007); Gaglianone, Lima, Linton and Smith (2011).

In this paper we focus on two aspects of risk measurement. First, risk measurement is closely related to utility theories of risk preferences. The papers that are most relevant to the present paper are Schmeidler (1986, 1989), which extend the expected utility theory by relaxing the independence axiom to the comonotonic independence axiom; this class of risk preference can successfully explain various violation of the expectation utility theory, such as the Ellsberg paradox.

Second, a major difficulty in measuring tail risk is that the tail part of a loss distribution is difficult to estimate and hence bears substantial model uncertainty. As emphasized by Hansen (2013), "uncertainty can come from limited data, unknown models and misspecification of those models," and "we can think of statistical models as approximations and we use such models in sophisticated ways with conservative adjustments that reflect the potential for misspecification."

In face of statistical uncertainty, different procedures may be used to forecast the risk measure. It is hence desirable to be able to evaluate which procedure gives a better forecast. The elicitability of a risk measure is a property based on a decision-theoretical framework for evaluating the performance of different forecasting procedures (Gneiting; 2011). The elicitability of a risk measure means that the risk measure can be obtained by minimizing the expectation of a forecasting objective function; hence, the forecasting objective function can then be used for evaluating different forecasting procedures.

Elicitability is closely related to backtesting, whose objective is to evaluate the

performance of a risk forecasting model. If a risk measure is elicitable, then the sample average forecasting error based on the objective function can be used for backtesting the risk measure. Gneiting (2011) shows that VaR is elicitable but expected shortfall is not, which "may challenge the use of the expected shortfall as a predictive measure of risk, and may provide a partial explanation for the lack of literature on the evaluation of expected shortfall forecasts, as opposed to quantile or VaR forecasts." Gaglianone, Lima, Linton and Smith (2011) propose a backtest for evaluating VaR estimates that delivers more power in finite samples than existing methods and develop a mechanism to find out why and when a model is misspecified; see also Jorion (2007, Ch. 6). Linton and Xiao (2013) show that VaR has an advantage over expected shortfall as the asymptotic inference procedures for VaR "has the same asymptotic behavior regardless of the thickness of the tails."

The main result of the paper is that the only tail risk measure that satisfies both a set of economic axioms proposed by Schmeidler (1989) and the statistical requirement of elicitability Gneiting (2011) is median shortfall, which is the median of the tail loss distribution and is also the VaR at a higher confidence level. In addition, we show that median shortfall has the desirable property of distributional robustness with respective to model misspecification in the sense of Hampel (1971), which means that a small deviation of the model only results in a small change in the risk measurement.

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 has distributional robustness. 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 models may be used; see Gilboa and Schmeidler (1989); Maccheroni, Marinacci and Rustichini (2006); Hansen and Sargent (2001, 2007); Gilboa, Maccheroni, Marinacci and Schmeidler (2010). We also incorporate multiple models in this paper; see Section 3. We complement these papers by studying (i) the link between risk measures and statistical uncertainty via ellicitability and (ii) distributional robustness of risk measures.

Important contribution to measurement of risk based on economic axioms includes Aumann and Serrano (2008) and Foster and Hart (2009, 2012), which study risk measurement of gambles (i.e., random variables with positive mean and take negative values with positive probability). This paper complement their results by linking economic axioms for risk measurement with statistical model uncertainty; in addition,

our approach focuses on the measurement of tail risk for general random variables. Thus, the risk measure considered in this paper has a different objective. As pointed out by Aumann and Serrano (2008), "like any index or summary statistic... the riskiness index summarizes a complex, high-dimensional object by a single number. Needless to say, no index captures all the relevant aspects of the situation being summarized."

The remainder of the paper is organized as follows. Section 2 presents the main result of the paper. In Section 3, we propose to use a scenario aggregation function to combine risk measurements under multiple models. In Section 4, we apply the results in previous sections to the study of Basel accord capital requirements. Section 5 discusses some comments related to median shortfall.

2 Main Results

2.1 Axioms and Representation

Let $(\Omega, \mathcal{F}, P)$ be a probability space that describes the states and the probability of occurrence of states at a future time T. Assume the probability space is large enough so that one can define a random variable uniformly distributed on [0,1]. Let a random variable X defined on the probability space denote the random loss of a portfolio of financial assets that will be realized at time T. Then $-X$ is the random profit of the portfolio. Let $\mathcal{X} \supset \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$ be a set of random variables. A risk measure ρ is a functional defined on $\mathcal{X}$ that maps a random variable X to a real number $\rho(X)$. The specification of $\mathcal{X}$ depends on ρ; in particular, $\mathcal{X}$ can include unbounded random variables. For example, if ρ is the variance, then $\mathcal{X}$ can be specified as $\mathcal{L}^2(\Omega, \mathcal{F}, P)$; if ρ is Value-at-Risk (VaR), then $\mathcal{X}$ can be specified as the set of all proper random variables.

An important relation between two random variables is a.s. comonotonicity (see, e.g., Schmeidler; 1986): Two random variables X and Y are said to be a.s. comonotonic, if there exists $\Omega_0 \subset \Omega$ such that $P(\Omega_0) = 1$ and $(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0$, $\forall \omega_1, \omega_2 \in \Omega_0$.[1]

Suppose that there is a representative agent in the economy and he or she prefers $-X$ to $-Y$. If the agent is risk averse, we may say that $-X$ is less risky than $-Y$.

[1]The notion of comonotonicity is generalized to the notion of dual comonotonicity by Cerreia-Vioglio, Maccheroni, Marinacci and Montrucchio (2013b).

Motivated by this, we propose the following set of axioms, which are based on the axioms for the Choquet expected utility (Schmeidler; 1989), for the risk measure ρ.

Axiom A1. Comonotonic independence: for all pairwise a.s. comonotonic random variables X, Y, Z and for all $\alpha \in (0,1)$, $\rho(X) < \rho(Y)$ implies that $\rho(\alpha X + (1-\alpha)Z) < \rho(\alpha Y + (1-\alpha)Z)$.

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

Axiom A3. Standardization: $\rho(x \cdot 1_\Omega) = sx$, for all $x \in \mathbb{R}$, where $s > 0$ is a constant.

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

Axiom A5. Continuity: $\lim_{M\to\infty} \rho(\min(\max(X, -M), M)) = \rho(X)$.

The first two axioms are the axioms for the Choquet expected utility risk preferences (Schmeidler; 1989); Axiom A3 with $s = 1$ is used in (Schmeidler; 1986). In general, the constant s in Axiom A3 can be related to the "countercyclical indexing" method proposed in Gordy and Howells (2006), where a time-varying multiplier s that increases during booms and decreases during recessions is used to dampen the procyclicality of capital requirements; see also Brunnermeier and Pedersen (2009); Brunnermeier, Crocket, Goodhart, Perssaud and Shin (2009); Adrian and Shin (2014). The fourth axiom is standard for a law invariant risk measure. The last continuity axiom states that the risk measurement of an unbounded random variable can be approximated by that of bounded random variables.

A function $h : [0,1] \to [0,1]$ is called a distortion function if $h(0) = 0$, $h(1) = 1$, and h is increasing. As a direct application of the results in (Schmeidler; 1986), we obtain the following representation of a risk measure that satisfies Axioms A1-A5.

Lemma 2.1. *Let $\mathcal{X} \supset \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$ be a set of random variables ($\mathcal{X}$ may include unbounded random variables). A risk measure $\rho : \mathcal{X} \to \mathbb{R}$ satisfies axioms A1-A5 if and only if there exists a distortion function $h(\cdot)$ such that*

$$\rho(X) = s \int X \, d(h \circ P) \tag{1}$$

$$= s \int_{-\infty}^{0} (h(P(X > x)) - 1) dx + s \int_{0}^{\infty} h(P(X > x)) dx, \ \forall X \in \mathcal{X}, \tag{2}$$

where the integral in (1) is the Choquet integral of X with respect to the distorted non-additive probability $h \circ P(A) := h(P(A))$, $\forall A \in \Sigma$.

Proof. Without loss of generality, we only need to prove for the case $s = 1$, as ρ satisfies Axioms A1-A5 if and only if $\frac{1}{s}\rho$ satisfies Axioms A1-A5 with $s = 1$.

The "only if" part. We will first show that (2) holds for all $X \in \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$. Define the set function $\nu(E) := \rho(1_E), E \in \mathcal{F}$. Then, it follows from Axiom A2 and A3 that ν is monotonic, $\nu(\emptyset) = 0$, and $\nu(\Omega) = 1$. By Axiom A2, $\rho(X) = \rho(Y)$ if $X = Y$ a.s. For $K \geq 1$, define $\mathcal{L}^M := \{X \mid |X| \leq M\}$ and define $\mathcal{L}^\infty := \cup_{M=1}^\infty \mathcal{L}^M$. For any $X \in \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$, let M_0 be the essential supremum of $|X|$ and denote $X^{M_0} := \min(M_0, \max(X, -M_0))$. Then $X^{M_0} \in \mathcal{L}^{M_0}$ and $X = X^{M_0}$ a.s., which implies that $\nu(X > x) = \nu(X^{M_0} > x)$, $\forall x$. Since ρ satisfies Axioms A1-A3 on $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$, it follows that ρ satisfies the conditions (i)-(iii) of the Corollary in Section 3 of Schmeidler (1986) (with $B(K)$ in the corollary defined to be $\mathcal{L}^{1+M_0}$). Hence, it follows from the Corollary that

$$\begin{aligned}\rho(X) = \rho(X^{M_0}) &= \int_0^\infty \nu(X^{M_0} > x)dx + \int_{-\infty}^0 (\nu(X^{M_0} > x) - 1)dx \\ &= \int_0^\infty \nu(X > x)dx + \int_{-\infty}^0 (\nu(X > x) - 1)dx. \end{aligned} \tag{3}$$

Define the function h such that $h(0) = 0$ and $h(1) = 1$. Let U be a uniform $U(0, 1)$ random variable. For any fixed $p \in (0, 1)$, define $h(p) := \rho(1_{\{U \leq p\}})$. By Axiom A4, $h(\cdot)$ satisfies $\nu(A) = h(P(A))$ for all A. Therefore, by (3), (2) holds for X. In addition, for any $0 < q < p < 1$, $h(p) = \rho(1_{\{U \leq p\}}) \geq \rho(1_{\{U \leq q\}}) = h(q)$. Hence, h is an increasing function.

Second, we show that (2) holds for any (possibly unbounded) $X \in \mathcal{X}$. For $M > 0$, since X^M belongs to $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$, it follows that (2) holds for X^M, which implies

$$\begin{aligned}\rho(X^M) &= \int_0^\infty h(P(X^M > x))dx + \int_{-\infty}^0 (h(P(X^M > x)) - 1)dx \\ &= \int_0^M h(P(X > x))dx + \int_{-M}^0 (h(P(X > x)) - 1)dx.\end{aligned}$$

Finally, letting $M \to \infty$ on both sides of the above equation and using Axiom A5, we conclude that (2) holds for X.

The "if" part. Suppose a risk measure ρ is defined by (2). Define the set function $\nu(A) := h(P(A)), \forall A \in \mathcal{F}$. Then ν is a monotonic set function and $\rho(X)$ is equal to the Choquet integral of X with respect to ν. It follows from properties of Choquet integral (Denneberg; 1994) that ρ satisfies Axioms A1-A5. $\square$

Lemma 2.1 extends the representation theorem in Wang, Young and Panjer (1997) as the requirement of $\lim_{d \to 0} \rho((X - d)^+) = \rho(X^+)$ in their continuity axiom is not

needed here.[2] Note that in the case of random variables, the main theorem in Schmeidler (1986) requires the random variables to be bounded, but Lemma 2.1 does not; Axiom A5 is automatically satisfied for bounded random variables.

It is clear from (2) that any risk measure satisfying Axioms A1-A5 is monotonic with respect to first-order stochastic dominance,[3] an important property of risk measurement emphasized by Aumann and Serrano (2008) and Foster and Hart (2009). Many commonly used risk measures are special cases of risk measures defined in (2).

Example 1. Value-at-Risk (VaR). VaR is a quantile of the loss distribution at some pre-defined probability level. More precisely, let X be the random loss with general distribution function $F_X(\cdot)$, which may not be continuous or strictly increasing. For a given $\alpha \in (0, 1]$, VaR of X at level α is defined as

$$\mathrm{VaR}_\alpha(X) := F_X^{-1}(\alpha) = \inf\{x \mid F_X(x) \geq \alpha\}.$$

For $\alpha = 0$, VaR of X at level α is defined to be $\mathrm{VaR}_0(X) := \inf\{x \mid F_X(x) > 0\}$ and $\mathrm{VaR}_0(X)$ is equal to the essential infimum of X. VaR is monotonic with respect to first-order stochastic dominance. Duffie and Pan (1997, 2001) and Jorion (2007), among others, provide comprehensive discussions of VaR and risk management. $\rho(X)$ in (2) is equal to $\mathrm{VaR}_\alpha(X)$ if $h(x) := 1_{\{x>1-\alpha\}}$.

Example 2. Expected shortfall (ES). For $\alpha \in (0, 1)$, ES of X at level α is defined as[4] the mean of the α-tail distribution of X (Tasche; 2002; Rockafellar and Uryasev;

[2] The axioms used in Wang, Young and Panjer (1997), including a comonotonic additivity axiom, imply Axioms A1-A5. More precisely, let $\mathbb{Q}$ and $\mathbb{Q}^+$ denote the set of rational numbers and positive rational numbers, respectively. Without loss of generality, suppose $s = 1$ in Axiom A3. (i) Their comonotonic additivity axiom implies that $\rho(\lambda X) = \lambda\rho(X)$ for any X and $\lambda \in \mathbb{Q}^+$, which in combination with their standardization axiom $\rho(1) = 1$ implies $\rho(\lambda) = \lambda\rho(1) = \lambda$, $\lambda \in \mathbb{Q}^+$. Since $\rho(-\lambda)+\rho(\lambda) = \rho(0) = 0$, it follows that $\rho(\lambda) = \lambda$, $\forall\lambda \in \mathbb{Q}$. Then for any $\lambda \in \mathbb{R}$, there exists $\{x_n\} \subset \mathbb{Q}$ and $\{y_n\} \subset \mathbb{Q}$ such that $x_n \downarrow \lambda$ and $y_n \uparrow \lambda$. By the monotonic axiom, $x_n = \rho(x_n) \geq \rho(\lambda) \geq \rho(y_n) = y_n$. Letting $n \to \infty$ yields $\rho(\lambda) = \lambda$, $\forall\lambda \in \mathbb{R}$; hence, Axiom A3 holds. (ii) By the monotonic axiom, $\rho(\min(X, M)) \leq \rho(\min(\max(X, -M), M)) \leq \rho(\max(X, -M))$. Letting $M \to \infty$ and using the conditions $\rho(\min(X, M)) \to \rho(X)$ and $\rho(\max(X, -M)) \to \rho(X)$ as $M \to \infty$ in their continuity axiom, *without need of the condition* $\lim_{d\to 0} \rho((X - d)^+) = \rho(X^+)$, Axiom A5 follows. (iii) We then show positive homogeneity holds, i.e. $\rho(\lambda X) = \lambda\rho(X)$ for any X and any $\lambda > 0$. For any X and $M > 0$, denote $X^M := \min(\max(X, -M), M)$. For any $\epsilon > 0$ and $\lambda > 0$, there exist $\{\lambda_n\} \subset \mathbb{Q}^+$ such that $\lambda_n \to \lambda$ as $n \to \infty$ and $\lambda_n\rho(X^M)-\epsilon = \rho(\lambda_n X^M-\epsilon) \leq \rho(\lambda X^M) \leq \rho(\lambda_n X^M+\epsilon) = \lambda_n\rho(X^M)+\epsilon$. Letting $n \to \infty$ yields $\lambda\rho(X^M) - \epsilon \leq \rho(\lambda X^M) \leq \lambda\rho(X^M) + \epsilon$, $\forall\epsilon > 0$. Letting $\epsilon \downarrow 0$ leads to $\rho(\lambda X^M) = \lambda\rho(X^M)$, $\forall\lambda \geq 0$. Letting $M \to \infty$ and applying Axiom A5 result in $\rho(\lambda X) = \lambda\rho(X)$, $\forall\lambda \geq 0$. Their comonotonic additivity axiom and positive homogeneity imply Axiom A1.

[3] For two random variables X and Y, if X first-order stochastically dominates Y, then $P(X > x) \geq P(Y > x)$ for all x, which implies that for a risk measure ρ represented in (2), $\rho(X) \geq \rho(Y)$.

[4] For $\alpha = 1$, ES of X at level α is defined as $\mathrm{ES}_1(X) := F_X^{-1}(1)$. For $\alpha = 0$, $F_{0,X} := F_X$ and ES of X at level 0 is $\mathrm{ES}_0(X) = E(X)$.

2002), i.e.,

$$\mathrm{ES}_\alpha(X) := \text{mean of the } \alpha\text{-tail distribution of } X = \int_{-\infty}^{\infty} x dF_{\alpha,X}(x),\ \alpha \in [0,1),$$

where $F_{\alpha,X}(x)$ is the α-tail distribution defined as (Rockafellar and Uryasev; 2002):

$$F_{\alpha,X}(x) := \begin{cases} 0, & \text{for } x < \mathrm{VaR}_\alpha(X) \\ \frac{F_X(x)-\alpha}{1-\alpha} & \text{for } x \geq \mathrm{VaR}_\alpha(X). \end{cases}$$

If the loss distribution F_X is continuous, then $F_{\alpha,X}$ is the same as the conditional distribution of X given that $X \geq \mathrm{VaR}_\alpha(X)$; if F_X is not continuous, then $F_{\alpha,X}(x)$ is a just a slight modification of the conditional loss distribution. For $\alpha \in [0,1)$, $\rho(X)$ in (2) is equal to $\mathrm{ES}_\alpha(X)$ if[5]

$$h(x) = \begin{cases} \frac{x}{1-\alpha}, & x \leq 1-\alpha, \\ 1, & x \geq 1-\alpha. \end{cases}$$

Example 3. Median shortfall (MS). As we will see later, expected shortfall has several statistical drawbacks including non-elicitability and non-robustness. To mitigate the problems, one may simply use median shortfall. In contrast to ES which is the mean of the tail loss distribution, MS is the median of the same tail loss distribution. More precisely, MS of X at level $\alpha \in [0,1)$ is defined as (Kou, Peng and Heyde; 2013)[6]

$$\begin{aligned} \mathrm{MS}_\alpha(X) &:= \text{median of the } \alpha\text{-tail distribution of } X \\ &= F^{-1}_{\alpha,X}(\frac{1}{2}) = \inf\{x \mid F_{\alpha,X}(x) \geq \frac{1}{2}\}. \end{aligned}$$

For $\alpha = 1$, MS at level α is defined as $\mathrm{MS}_1(X) := F_X^{-1}(1)$. Therefore, MS at level α can *capture the tail risk beyond the VaR at level* α, because it measures the median of the loss size conditional on that the loss exceeds the VaR at level α. It can be shown that[7]

$$\mathrm{MS}_\alpha(X) = \mathrm{VaR}_{\frac{1+\alpha}{2}}(X),\ \forall X,\ \forall \alpha \in [0,1].$$

[5] $\rho(X)$ in (2) is equal to $\mathrm{ES}_1(X)$ if $h(x) = 1_{\{x>0\}}$.

[6] The term "median shortfall" is also used in Moscadelli (2004) and So and Wong (2012) but is respectively defined as median$[X|X > u]$ for a constant u and median$[X|X > \mathrm{VaR}_\alpha(X)]$, which are different from ours.

[7] Indeed, for $\alpha \in (0,1)$, by definition, $\mathrm{MS}_\alpha(X) = \inf\{x \mid F_{\alpha,X}(x) \geq \frac{1}{2}\} = \inf\{x \mid \frac{F_X(x)-\alpha}{1-\alpha} \geq \frac{1}{2}\} = \inf\{x \mid F_X(x) \geq \frac{1+\alpha}{2}\} = \mathrm{VaR}_{\frac{1+\alpha}{2}}(X)$; for $\alpha = 1$, by definition, $\mathrm{MS}_1(X) = F_X^{-1}(1) = \mathrm{VaR}_1(X)$; for $\alpha = 0$, by definition, $F_{0,X} = F_X$ and hence $\mathrm{MS}_0(X) = F_X^{-1}(\frac{1}{2}) = \mathrm{VaR}_{\frac{1}{2}}(X)$.

Hence, $\rho(X)$ in (2) is equal to $\mathrm{MS}_\alpha(X)$ if $h(x) := 1_{\{x>(1-\alpha)/2\}}$.

Example 4. Generalized spectral risk measures. A generalized spectral risk measure is defined by

$$\rho_m(X) := \int_{(0,1]} F_X^{-1}(u)dm(u), \tag{4}$$

where m is a probability measure on $(0,1]$. The class of risk measures represented by (2) include and are strictly larger than the class of generalized spectral risk measures, as they all satisfy Axioms A1-A5.[8] A special case of (4) is the spectral risk measure, defined as (Acerbi; 2002)

$$\rho(X) = \int_{[0,1]} \mathrm{ES}_u(X)d\tilde{m}(u),$$

where $\tilde{m}$ is a probability measure on $[0,1]$, by choosing $\frac{dm(u)}{du} = \int_{[0,u)} \frac{1}{1-y} d\tilde{m}(y)$ for $u \in (0,1)$ and $m(\{1\}) = \tilde{m}(\{1\})$. The MINMAXVAR risk measure proposed in Cherny and Madan (2009) for the measurement of trading performance is a special case of the spectral risk measure, corresponding to a distortion function $h(x) = 1-(1-x^{\frac{1}{1+\alpha}})^{1+\alpha}$ in (2), where $\alpha \geq 0$ is a constant.

2.2 Elicitability

In practice, the measurement of the risk of X using ρ is a point forecasting problem, because the true distribution F_X is unknown and one has to find an estimate $\hat{F}_X$ and use $\rho(\hat{F}_X)$ to forecast the unknown true value of $\rho(F_X)$. As one may come up with different point forecasts $\rho(\hat{F}_X)$ based on different $\hat{F}_X$, it is an important issue to evaluate which point forecast provides a better forecast of $\rho(F_X)$.

Elicitability is a recently developed theory that lays the decision-theoretical foundation for effective evaluation of point forecasting procedures. Suppose one wants to

[8] In fact, for any fixed $u \in (0,1]$, it is well-known that $F_X^{-1}(u)$ as a functional on $\mathcal{L}^\infty(\Omega,\mathcal{F},P)$ satisfies positive homogeneity and comonotonic additivity, which implies that ρ_m satisfies Axioms A1-A4. On $\mathcal{L}^\infty(\Omega,\mathcal{F},P)$, ρ_m apparently satisfies Axiom A5. On the other hand, for an $\alpha \in (0,1)$, the right quantile $q_\alpha^+(X) := \inf\{x \mid F_X(x) > \alpha\}$ is a special case of the risk measure defined in (2) with $h(x)$ being defined as $h(x) := 1_{\{x \geq 1-\alpha\}}$, but cannot be represented by (4). Indeed, we will show that there does not exist m such that $q_\alpha^+(X) = \rho_m(X)$, $\forall X \in \mathcal{L}^\infty(\Omega,\mathcal{F},P)$. Suppose for the sake of contradiction that such an m exists. Let $X_0 \in \mathcal{L}^\infty(\Omega,\mathcal{F},P)$ have a strictly positive density on its support. Then, $F_{X_0}^{-1}(u)$ as a function of u strictly increases on $(0,1]$. Hence, $q_\alpha^+(X_0) = \rho_m(X_0)$ implies that m is the point mass on α, which leads to $q_\alpha^+(X) = \rho_m(X) = F_X^{-1}(\alpha)$, $\forall X \in \mathcal{L}^\infty(\Omega,\mathcal{F},P)$. However, this equality does not hold for X with $q_\alpha^+(X) > F_X^{-1}(\alpha)$.

forecast the realization of a random variable Y using a point x, without knowing the true distribution F_Y of Y. The forecasting error can be measured by

$$ES(x, Y) = \int S(x, y) dF_Y(y),$$

where $S(x, y) : \mathbb{R}^2 \to \mathbb{R}$ is a forecasting objective function, e.g., $S(x, y) = (x-y)^2$ and $S(x, y) = |x - y|$. As F_Y is unknown, the forecasting error can be approximated by the average $\frac{1}{n}\sum_{i=1}^{n} S(x_i, Y_i)$, where $Y_1, \ldots, Y_n$ are samples of Y and $x_1, \ldots, x_n$ are the corresponding point forecasts. Then two point forecasting methods can be evaluated by comparing their average forecasting errors. The optimal forecast corresponding to a given objective function S is

$$\rho^*(F_Y) = \arg\min_x ES(x, Y).$$

For example, when $S(x, y) = (x - y)^2$ and $S(x, y) = |x - y|$, the optimal forecast is the mean functional $\rho^*(F_Y) = E(Y)$ and the median functional $\rho^*(F_Y) = F_Y^{-1}(\frac{1}{2})$, respectively.

A statistical functional ρ is elicitable if there exists a forecasting objective function S such that minimizing the expected forecasting error yields ρ. Many statistical functionals are elicitable. For example, minimizing the forecasting error with $S(x, y) = |x - y|$ yields the median functional; hence, the median functional is elicitable. If a statistical functional is not elicitable, then one cannot find such a forecasting objective function. In this case, for any objective function S, the minimization of the forecasting error does not yield the true value $\rho(F)$. Hence, one cannot tell which one of competing point forecasts for $\rho(F)$ performs the best by comparing their forecasting error, no matter what objective function S is used.

The general concept of elicitability dates back to the pioneering work of Savage (1971); Thomson (1979); Osband (1985) and is comprehensively developed in Gneiting (2011). Gneiting (2011) contends that “in issuing and evaluating point forecasts, it is essential that either the objective function (i.e., $S(\cdot)$) be specified ex ante, or an elicitable target functional be named, such as an expectation or a quantile, and objective functions be used that are consistent for the target functional.” Engelberg, Manski and Williams (2009) also points out the critical importance of the specification of an objective function or an elicitable target functional. The importance of elicitability of a risk measure is discussed in Embrechts and Hofert (2014).

In the present paper, we are concerned with the measurement of risk, which is single-valued statistical functional. Following Definition 2 in Gneiting (2011), which defines elicitability for a set-valued statistical functional, we define the elicitability for a single-valued statistical functional as follows.[9]

Definition 2.1. *A single-valued statistical functional $\rho(\cdot)$ is elicitable with respect to a class of distributions $\mathcal{P}$ if there exists a forecasting objective function $S : \mathbb{R}^2 \to \mathbb{R}$ such that*

$$\rho(F) = \min \left\{ x \mid x \in \arg\min_x \int S(x, y) dF(y) \right\}, \ \forall F \in \mathcal{P}. \tag{5}$$

2.3 Main Result

The following Theorem 2.1 shows that median shortfall is the *only* risk measure that (i) captures tail risk; (ii) is elicitable; and (iii) has the decision theoretical foundation of Choquet expected utility, because the mean apparently does not capture tail risk.

Theorem 2.1. *Let $\rho(\cdot)$ be a risk measure defined on $\mathcal{X} \supset \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$ that satisfies Axioms A1-A5. Let $\mathcal{P} = \{F_X \mid X \in \mathcal{X}\}$. Then, $\rho(\cdot)$ is elicitable with respect to $\mathcal{P}$ if and only if one of the following two cases holds:*

(i) $\rho = VaR_\alpha$ for some $\alpha \in (0, 1]$ (noting that $MS_\alpha = VaR_{\frac{\alpha+1}{2}}$ for $\alpha \in [0, 1]$).

(ii) $\rho(F) = \int x dF(x)$, $\forall F$.

Proof. See Appendix A. □

The major difficulty of the proof lies in that the distortion function $h(\cdot)$ in the representation equation (2) of risk measures satisfying Axioms A1-A5 can have various kinds of discoutinuities on $[0, 1]$. The outline of the proof is as follows. First, we show that the necessary condition for ρ to be elicitable is that ρ has convex level sets, i.e., $\rho(F_1) = \rho(F_2)$ implies that $\rho(F_1) = \rho(\lambda F_1 + (1 - \lambda)F_2)$, $\forall \lambda \in (0, 1)$. The second and the key step is to show that only three kinds of risk measures have convex level sets: (i) VaR_α, $\alpha \in (0, 1]$, and, in particular, MS_α, $\alpha \in [0, 1]$; (ii) the mean functional $\rho(F) = mean(F)$; (iii) $\rho(F) = cq_\alpha^-(F) + (1 - c)q_\alpha^+(F)$, where $c \in (0, 1)$ is a constant,

[9] The requirement that $\rho(F)$ is the *minimum* of the set of minimizers of the expected objective function is not essential. In fact, if one replaces the first "min" in (5) by "max", the conclusions of the paper remain the same; one only needs to change "VaR_α" to the right quantile q_α^+ in Theorem 2.1.

$q^-_\alpha(F) := \inf\{x \mid F(x) \geq \alpha\}$, and $q^+_\alpha(F) := \inf\{x \mid F(x) > \alpha\}$. Lastly, we show that $\rho(F) = cq^-_\alpha(F) + (1-c)q^+_\alpha(F)$ is not elicitable by extending the main proposition in Thomson (1979).

2.4 Statistical Robustness of MS

One commonly accepted definition of robustness is given explicitly in (Hampel; 1971). Suppose the true loss distribution is F and we want to calculate $\rho(F)$. However, since the true distribution F is unknown, we have to use a model $\hat{F}$ to approximate F and what we can compute is $\rho(\hat{F})$ instead of $\rho(F)$. ρ is insensitive to model misspecification means that: if the misspecified model $\hat{F}$ only deviates a little from the true F, then $\rho(\hat{F})$ only deviates a little from $\rho(F)$.

Let $\mathcal{M}$ be a set of probability measures. Hampel's theorem (see Huber and Ronchetti; 2009, Section 2.6) shows that for a statistical functional $T(\cdot) : \mathcal{M} \to \mathbb{R}$, the sequence of statistics $T(F_n)$ is robust at F_0 if and only if T is continuous at F_0. It can be shown that MS is Hampel-robust but ES is not.

Lemma 2.2. *(i) For any $\alpha \in (0,1)$ and any F such that F^{-1} is continuous at $(1+\alpha)/2$, MS_α is Hampel-robust at F. (ii) For any $\alpha \in (0,1)$ and any F, ES_α is not Hampel-robust at F.*

Proof. Let $\delta_{(1+\alpha)/2}$ be the Dirac Delta measure on $(0,1)$ that has a point mass at $(1+\alpha)/2$ and denote $m(u) := \frac{u-\alpha}{1-\alpha}1_{\{u\geq\alpha\}}$, $u \in [0,1]$. Then MS_α and ES_α can be represented by $\text{MS}_\alpha(F) = \int_0^1 F^{-1}(u)d\delta_{(1+\alpha)/2}(u)$ and $\text{ES}_\alpha(F) = \int_0^1 F^{-1}(u)dm(u), \forall F$. The conclusion follows by applying Theorem 3.7 in Huber and Ronchetti (2009). $\square$

Kou, Peng and Heyde (2013) show that robustness is indispensable for external risk measures that are used for legal enforcement, such as risk measures for calculating trading book capital requirements, so that law can be enforced consistently, not sensitive to uncertainty in distribution.[10]

[10]Another meaning of robustness refers to that a small change of the data set, such as changing a few samples, or adding a few outliers to the data set, or making small changes to many samples, only results in a small change to the estimated risk measure (Huber and Ronchetti; 2009). Kou, Peng and Heyde (2013, Appendix F) also show that MS is a robust statistic but ES is not, by using another three tools of robust statistics, i.e, influence functions, asymptotic breakdown points, and finite-sample breakdown points; see also Cont, Deguest and Scandolo (2010).

3 Extension to Incorporate Multiple Models

The previous sections address the issue of model uncertainty from the perspective of elicitability and robustness. Following Gilboa and Schmeidler (1989) and Hansen and Sargent (2001), we further address the issue by considering multiple models (scenarios).[11] More precisely, we consider m probability measures P_i, $i = 1, \ldots, m$ on the state space (Ω, Σ). Each P_i corresponds to one model or one scenario, which may refer to a specific economic regime such as an economic boom and a financial crisis. The loss distribution $F_i(x) := P_i(X \leq x)$ of a random loss X under different scenarios can be substantially different. For example, the VaR calculated under the scenario of the 2007 financial crisis is much higher than that under a scenario corresponding to a normal market condition due to the difference of loss distributions.

Suppose that under the i-th scenario, the measurement of risk is given by $\rho_i(X) := \int X d(h_i \circ P_i)$, where h_i is a distortion function, $i = 1, \ldots, m$. We propose the following risk measure to incorporate multiple scenarios:

$$\rho(X) = f(\rho_1(X), \rho_2(X), \ldots, \rho_m(X)), \tag{6}$$

where $f : \mathbb{R}^m \to \mathbb{R}$ is a scenario aggregation function.

We postulate that the scenario aggregation function f satisfies the following axioms:

Axiom B1. Positive homogeneity and translation scaling: $f(a\tilde{x} + b\mathbf{1}) = af(\tilde{x}) + sb$, $\forall \tilde{x} \in \mathbb{R}^m, \forall a \geq 0, \forall b \in \mathbb{R}$, where $s > 0$ is a constant and $\mathbf{1} := (1, 1, ..., 1) \in \mathbb{R}^m$.

Axiom B2. Monotonicity: $f(\tilde{x}) \leq f(\tilde{y})$, if $\tilde{x} \leq \tilde{y}$, where $\tilde{x} \leq \tilde{y}$ means $x_i \leq y_i, i = 1, \ldots, m$.

Axiom B3. Uncertainty aversion: if $f(\tilde{x}) = f(\tilde{y})$, then for any $\alpha \in (0, 1)$, $f(\alpha\tilde{x} + (1 - \alpha)\tilde{y}) \leq f(\tilde{x})$.

Axiom B1 states that if the risk measurement of Y is an affine function of that of X under each scenario, then the aggregate risk measurement of Y is also a affine function of that of X. Axiom B2 is apparently the minimum requirement. Axiom B3 is proposed by Gilboa and Schmeidler (1989) to "capture the phenomenon of hedging"; it is used as one of the axioms for the maxmin expected utility that incorporates model uncertainty.

[11] Cerreia-Vioglio, Maccheroni, Marinacci and Montrucchio (2013a) show that the maxmin approach of Gilboa and Schmeidler (1989) is equivalent to the minimax approach in robust statistics.

Lemma 3.1. *A scenario aggregation function* $f : \mathbb{R}^m \to \mathbb{R}$ *satisfies Axioms B1-B3 if and only if there exists a set of weights* $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^m$ *with each* $\tilde{w} = (w_1, \ldots, w_m) \in \mathcal{W}$ *satisfying* $w_i \geq 0$ *and* $\sum_{i=1}^m w_i = 1$*, such that*

$$f(x) = s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m w_i x_i \right\}. \tag{7}$$

Proof. First, we show that Axioms B1-B3 are equivalent to the Axioms C1-C4 in Kou, Peng and Heyde (2013) with $n_i = 1$, $i = 1, \ldots, m$. Axioms B1 and B2 are the same as the Axioms C1 and C2, respectively. Axiom C4 holds for any function when $n_i = 1$, $i = 1, \ldots, m$. Axiom C3 and C1 apparently implies Axiom B3. We will then show that Axiom B3 and B1 implies Axiom C3. In fact, For any $\tilde{x}$ and $\tilde{y}$, it follows from Axiom B1 that $f(\tilde{x} - f(\tilde{x})/s) = f(\tilde{y} - f(\tilde{y})/s) = 0$. Then, it follows from Axiom B3 and Axiom B1 that $f(\tilde{x} + \tilde{y}) - f(\tilde{x}) - f(\tilde{y}) = f(\tilde{x} - f(\tilde{x})/s + \tilde{y} - f(\tilde{y})/s) = 2f(\frac{1}{2}(\tilde{x} - f(\tilde{x})/s) + \frac{1}{2}(\tilde{y} - f(\tilde{y})/s)) \leq 2f(\tilde{x} - f(\tilde{x})/s) = 0$. Hence, Axiom C3 holds. Therefore, Axioms B1-B3 are equivalent to Axioms C1-C4, and hence the conclusion of the lemma follows from Theorem 3.1 in Kou, Peng and Heyde (2013). □

In the representation (7), each weight $\tilde{w} \in \mathcal{W}$ can be regarded as a prior probability on the set of the scenarios; more precisely, w_i can be viewed as the likelihood that the scenario i happens.

Lemma 2.1 and Lemma 3.1 lead to the following class of risk measures:[12]

$$\rho(X) = s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m w_i \int X \, d(h_i \circ P_i) \right\}. \tag{8}$$

By Theorem 2, the requirement of elicitability under each scenario leads to the following tail risk measure

$$\rho(X) = s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m w_i \mathrm{MS}_{i,\alpha_i}(X) \right\}, \tag{9}$$

where $\mathrm{MS}_{i,\alpha_i}(X)$ is the median shortfall of X at confidence level α_i calculated under the i-th scenario (model). The risk measure ρ in (9) addresses the issue of model uncertainty from two aspects: (i) under each scenario i, MS_{i,α_i} is elicitable and robust; (ii) ρ incorporates multiple scenarios and multiple priors on the set of scenarios.

[12] Gilboa and Schmeidler (1989) consider $\inf_{P \in \mathcal{P}} \int u(X) \, dP$ without h_i; see also Xia (2013).

4 Application to Basel Accord Capital Rule for Trading Book

What risk measure should be used for setting capital requirement for banks is an important issue that has been under debate since the 2007 financial crisis. The Basel II use a 99.9% VaR for setting capital requirements for banking books of financial institutions (Gordy; 2003). The Basel II capital charge for the trading book on the t-th day is defined as $\rho(X) := s \max\left\{\frac{1}{s}\mathrm{VaR}_{t-1}(X), \frac{1}{60}\sum_{i=1}^{60}\mathrm{VaR}_{t-i}(X)\right\}$, where X is the trading book loss; $s \geq 3$ is a constant; $\mathrm{VaR}_{t-i}(X)$ is the 10-day VaR at 99% confidence level calculated on day $t-i$, which corresponds to the i-th model, $i = 1, \ldots, 60$. Define the 61-th model under which $P(X = 0) = 1$. Then, the Basel II risk measure is a special case of the class of risk measures considered in (9); it incorporates 61 models and two priors: one is $\tilde{w} = (1/s, 0, \ldots, 0, 1 - 1/s)$, the other $\tilde{w} = (1/60, 1/60, \ldots, 1/60, 0)$. The Basel 2.5 risk measure (Basel Committee on Banking Supervision; 2009) mitigates the procycliclity of the Basel II risk measure by incorporating the "stressed VaR" calculated under stressed market conditions such as financial crisis. The Basel 2.5 risk measure can also be written in the form of (9).

In a consultative document released by the Bank for International Settlement (Basel Committee on Banking Supervision; 2013), the Basel Committee proposes to "move from value-at-risk to expected shortfall," which "measures the riskiness of a position by considering both the size and the likelihood of losses above a certain confidence level." The proposed new Basel (called Basel 3.5) capital charge for the trading book measured on the t-th day is defined as $\rho(X) := s \max\left\{\frac{1}{s}\mathrm{ES}_{t-1}, \frac{1}{60}\sum_{i=1}^{60}\mathrm{ES}_{t-i}\right\}$, where ES_{t-i} is the ES at 97.5% confidence level calculated on day $t-i$, $i = 1, \ldots, 60$; hence, the Basel 3.5 risk measure is a special case of the class of risk measures considered in (8).[13]

The major argument for the change from VaR to ES is that a 99% VaR number of 100 million dollar does no carry information as to the size of loss in cases when the loss does exceed 100 million; on the other hand, the ES captures the tail risk in the sense that the 99% ES measures the mean of the size of loss given that the loss exceeds the 99% VaR.

[13]The Basel II, Basel 2.5, and newly proposed risk measure (Basel 3.5) for the trading book are all special cases of the class of risk measures called natural risk statistics proposed by Kou, Peng and Heyde (2013). The natural risk statistics are axiomatized by a different set of axioms including a comonotonic subadditivity axiom.

Although the argument sounds reasonable, the ES is not the only risk measure that captures the tail risk; in particular, an alternative risk measure that captures the tail risk is median shortfall, which, in contrast to expected shortfall, measures the median rather than the mean of the conditional tail distribution. For instance, in the aforementioned example, if we want to capture the tail risk, i.e., the size of the loss beyond the 99% VaR level, we can use either ES at 99% level or, alternatively, median shortfall at 99% level, which is the median of the same conditional loss distribution.

It follows from Theorem 2.1 and discussion in Section 2.4 that MS may be preferable than ES for setting capital requirements in banking regulation because MS is elicitable and robust, but ES is neither elicitable nor robust.

To further compare the robustness of MS with ES, we carry out a simple empirical study on the measurement of the tail risk of S&P 500 daily return. We consider two IGARCH(1, 1) models similar to the model of RiskMetrics:

- Model 1: IGARCH(1, 1) with conditional distribution being Gaussian

$$\begin{aligned} & r_t = \mu + \sigma_t \epsilon_t, \ \sigma_t^2 = \beta \sigma_{t-1}^2 + (1-\beta) r_{t-1}^2 \\ & \epsilon_t \overset{d}{\sim} N(0,1) \end{aligned}$$

- Model 2: the same as model 1 except that the conditional distribution is specified as $\epsilon_t \overset{d}{\sim} t_\nu$, where t_ν denotes t distribution with degree of freedom ν.

We respectively fit the two models to the historical data of daily returns of S&P 500 Index during 1/2/1980–11/26/2012 and then forecast the one-day MS and ES of a portfolio of S&P500 stocks that is worth 1,000,000 dollars on 11/26/2012. The comparison of the forecasts of MS and ES under the two models is shown in Table 1, where $\mathrm{ES}_{\alpha,i}$ and $\mathrm{MS}_{\alpha,i}$ are the ES_α and MS_α calculated under the i-th model, respectively, $i = 1, 2$. It is clear from the table that the change of ES under the two models (i.e., $\mathrm{ES}_{\alpha,2} - \mathrm{ES}_{\alpha,1}$) is much larger than that of MS (i.e., $\mathrm{MS}_{\alpha,2} - \mathrm{MS}_{\alpha,1}$), indicating that ES is more sensitive to model misspecification than MS.

5 Comments

It is worth noting that it is not desirable for a risk measure to be too sensitive to the tail risk. For example, let L denote the loss that could occur to a person who walks on the street. There is a very small but positive probability that the person could

Table 1: The comparison of the forecasts of one-day MS and ES of a portfolio of S&P500 stocks that is worth 1,000,000 dollars on 11/26/2012. $\mathrm{ES}_{\alpha,i}$ and $\mathrm{MS}_{\alpha,i}$ are the ES and MS at level α calculated under the i-th model, respectively, $i = 1, 2$.

α	ES			MS			$\frac{\mathrm{ES}_{\alpha,2}-\mathrm{ES}_{\alpha,1}}{\mathrm{MS}_{\alpha,2}-\mathrm{MS}_{\alpha,1}} - 1$
	$\mathrm{ES}_{\alpha,1}$	$\mathrm{ES}_{\alpha,2}$	$\mathrm{ES}_{\alpha,2} - \mathrm{ES}_{\alpha,1}$	$\mathrm{MS}_{\alpha,1}$	$\mathrm{MS}_{\alpha,2}$	$\mathrm{MS}_{\alpha,2} - \mathrm{MS}_{\alpha,1}$	
97.0%	19956	21699	1743	19070	19868	798	118.4%
97.5%	20586	22690	2104	19715	20826	1111	89.3%
98.0%	21337	23918	2581	20483	22011	1529	68.8%
98.5%	22275	25530	3254	21441	23564	2123	53.3%
99.0%	23546	27863	4317	22738	25807	3070	40.6%
99.5%	25595	32049	6454	24827	29823	4996	29.2%

be hit by a car and lose his life; in that unfortunate case, L may be infinite. Hence, the ES of L may be equal to infinity, suggesting that the person should never walk on the street, which is apparently not reasonable. In contrast, the MS of L is a finite number.

Theorem 2.1 generalizes the main result in Ziegel (2013), which shows the only elicitable spectral risk measure is the mean functional; note that VaR is not a spectral risk measure. Bellini and Bignozzi (2013) suggest a more restrictive definition of elicitability than Gneiting (2011); under their definition, median or quantile may not be elicitable, while they are always elicitable in the sense of Gneiting (2011). The elicitability of a risk measure is related to the concept of "consistency" of a risk measure introduced in (Davis; 2013).

The axioms in this paper are based on economic considerations. Other axioms based on mathematical considerations include subadditivity (Huber; 1981; Artzner et al.; 1999),[14] comonotonic subadditivity (Song and Yan; 2006; Kou, Peng and Heyde; 2013), convexity (Föllmer and Schied; 2002; Frittelli and Gianin; 2002), comonotonic convexity (Song and Yan; 2009).

The subadditivity axiom is somewhat controversial:[15] (i) The subadditivity axiom

[14] The representation theorem in (Artzner et al.; 1999) is based on Huber (1981), who use the same set of axioms. Gilboa and Schmeidler (1989) obtains a more general representation based on a different set of axioms.

[15] Even if one believes in subadditivity, VaR (and median shortfall) satisfies subadditivity in most relevant situations. In fact, Daníelsson, Jorgensen, Samorodnitsky, Sarma and de Vries (2013) show that VaR (and median shortfall) is subadditive in the relevant tail region if asset returns are regularly varying and possibly dependent, although VaR does not satisfies global subadditivity. Ibragimov and Walden (2007) and Ibragimov (2009) show that VaR is subadditive for the infinite variance stable

is based on an intuition that "a merger does not create extra risk" (Artzner et al.; 1999, p. 209), which may not be true, as can be seen from the merger of Bank of America and Merrill Lynch in 2008. (ii) Subadditivity is related to the idea that diversification is beneficial; however, diversification may not always be beneficial. Fama and Miller (1972, pp. 271–272) show that diversification is ineffective for asset returns with heavy tails (with tail index less than 1); these results are extended in Ibragimov and Walden (2007) and Ibragimov (2009). See Kou, Peng and Heyde (2013, Sec. 6.1) for more discussion. (iii) Although subadditivity ensures that $\rho(X_1)+\rho(X_2)$ is an upper bound for $\rho(X_1+X_2)$, this upper bound may not be valid in face of model uncertainty.[16] (iv) In practice, $\rho(X_1)+\rho(X_2)$ may not be a useful upper bound for $\rho(X_1+X_2)$ as the former may be too larger than the latter.[17] (v) Subadditivity is not necessarily needed for capital allocation or asset allocation.[18]

A Proof of Theorem 2.1

First, we give the following definition:[19]

Definition A.1. *A single-valued statistical functional ρ is said to have convex level sets with respect to $\mathcal{P}$, if for any two distributions $F_1 \in \mathcal{P}$ and $F_2 \in \mathcal{P}$, $\rho(F_1) = \rho(F_2)$ implies that $\rho(\lambda F_1 + (1-\lambda)F_2) = \rho(F_1)$, $\forall \lambda \in (0,1)$.*

distributions with finite mean. "In this sense, they showed that VaR is subadditive for the tails of all fat distributions, provided the tails are not super fat (e.g., Cauthy distribution)" (Gaglianone, Lima, Linton and Smith; 2011).

[16]In fact, suppose we are concerned with obtaining an upper bound for $\mathrm{ES}_\alpha(X_1+X_2)$. In practice, due to model uncertainty, we can only compute $\widehat{ES(X_1)}$ and $\widehat{ES(X_2)}$, which are estimates of $ES(X_1)$ and $ES(X_2)$ respectively. $\widehat{ES(X_1)}+\widehat{ES(X_2)}$ cannot be used as an upper bound for $ES(X_1+X_2)$ because it is possible that $\widehat{ES(X_1)}+\widehat{ES(X_2)} < ES(X_1)+ES(X_2)$.

[17]For example, let X_1 be the loss of a long position of a call option on a stock (whose price is \$100) at strike \$100 and let X_2 be the loss of a short position of a call option on that stock at strick \$95. Then the margin requirement for X_1+X_2, $\rho(X_1+X_2)$, should not be larger than \$5, as $X_1+X_2 \le 5$. However, $\rho(X_1)=0$ and $\rho(X_2)\approx 20$ (the margin is around 20% of the underlying stock price). In this case, no one would use the subadditivity to charge the upper bound $\rho(X_1)+\rho(X_2)\approx 20$ for the portfolio X_1+X_2; instead, people will directly compute $\rho(X_1+X_2)$.

[18]Kou et al. (2013, Sec. 7) derive the Euler capital allocation rule for a class of risk measures including VaR with scenario analysis and the Basel Accord risk measures; see Wen et al. (2013), Xi et al. (2013), and the references therein for asset allocation methods using VaR and Basel Accord risk measures.

[19]A similar definition for a set-valued (not single-valued) statistical functional is given in Osband (1985); Gneiting (2011).

The following Lemma A.1 gives a necessary condition for a single-valued statistical functional to be elicitable.

Lemma A.1. *If a single-valued statistical functional ρ is elicitable with respect to $\mathcal{P}$, then ρ has convex level sets with respect to $\mathcal{P}$.*

Proof. Suppose ρ is elicitable. Then there exists a forecasting objective function $S(x,y)$ such that (5) holds. For any two distribution F_1 and F_2 and any $\lambda \in (0,1)$, denote $F_\lambda = \lambda F_1 + (1-\lambda)F_2$. If $t = \rho(F_1) = \rho(F_2)$, then $t = \min\{x \mid x \in \arg\min_x \int S(x,y)dF_i(y)\}$, $i = 1,2$. Since $\int S(x,y)dF_\lambda(y) = \lambda \int S(x,y)dF_1(y) + (1-\lambda)\int S(x,y)dF_2(y)$, it follows that $t \in \arg\min_x \int S(x,y)dF_\lambda(y)$. For any $t' \in \arg\min_x \int S(x,y)dF_\lambda(y)$, it holds that $\int S(t',y)dF_\lambda(y) \leq \int S(t,y)dF_\lambda(y)$, which implies that $\lambda \int S(t',y)dF_1(y) + (1-\lambda)\int S(t',y)dF_2(y) \leq \lambda \int S(t,y)dF_1(y) + (1-\lambda)\int S(t,y)dF_2(y)$. However, by definition of t, $\int S(t,y)dF_i(y) \leq \int S(t',y)dF_i(y), i = 1,2$. Therefore, $\int S(t,y)dF_i(y) = \int S(t',y)dF_i(y), i = 1,2$, which implies that $t' \in \arg\min_x \int S(x,y)dF_i(y)$, $i = 1,2$. Since $t = \min\{x \mid x \in \arg\min_x \int S(x,y)dF_i(y)\}$, it follows that $t' \geq t$. Therefore, $t = \min\{x \mid x \in \arg\min_x \int S(x,y)dF_\lambda(y)\} = \rho(F_\lambda)$. $\square$

Lemma A.2. *Let $\alpha \in (0,1)$ and $c \in [0,1)$. Let ρ be defined in (2) with h being defined as*

$$h(x) := \begin{cases} 0, & \text{if } x < 1-\alpha, \\ 1-c, & \text{if } x = 1-\alpha, \\ 1, & \text{if } x > 1-\alpha. \end{cases}$$

Then

$$\rho(F) = cq_\alpha^-(F) + (1-c)q_\alpha^+(F),\ \forall F \in \mathcal{P}, \tag{10}$$

where $q_\alpha^-(F) := \inf\{x \mid F(x) \geq \alpha\}$ and $q_\alpha^+(F) := \inf\{x \mid F(x) > \alpha\}$. Furthermore, ρ has convex level sets with respect to $\mathcal{P}$.

Proof. Define $g(x) := 1 - h(1-x)$, $x \in [0,1]$. Then,

$$g(x) = \begin{cases} 0, & \text{if } x < \alpha, \\ c, & \text{if } x = \alpha, \\ 1, & \text{if } x > \alpha, \end{cases}$$

and ρ can be represented as

$$\rho(F) = -\int_{-\infty}^0 g(F(x))dx + \int_0^\infty (1 - g(F(x)))dx.$$

Consider three cases:

(i) $q_\alpha^-(F) \geq 0$. In this case,

$$\begin{aligned}\rho(F) &= \int_0^\infty (1-g(F(x)))dx \\ &= \int_{[0,q_\alpha^-(F))} (1-g(F(x)))dx + \int_{[q_\alpha^-(F),q_\alpha^+(F))} (1-g(F(x)))dx \\ &\quad + \int_{(q_\alpha^+(F),\infty)} (1-g(F(x)))dx \\ &= q_\alpha^-(F) + (1-c)(q_\alpha^+(F) - q_\alpha^-(F)) \\ &= cq_\alpha^-(F) + (1-c)q_\alpha^+(F).\end{aligned}$$

(ii) $q_\alpha^-(F) < 0 < q_\alpha^+(F)$. In this case,

$$\begin{aligned}\rho(F) &= -\int_{(q_\alpha^-(F),0)} g(F(x))dx + \int_{(0,q_\alpha^+(F))} (1-g(F(x)))dx \\ &= cq_\alpha^-(F) + (1-c)q_\alpha^+(F).\end{aligned}$$

(iii) $q_\alpha^+(F) \leq 0$. In this case,

$$\begin{aligned}\rho(F) &= -\int_{(-\infty,q_\alpha^-(F))} g(F(x))dx - \int_{(q_\alpha^-(F),q_\alpha^+(F))} g(F(x))dx - \int_{(q_\alpha^+(F),0)} g(F(x))dx \\ &= -c(q_\alpha^+(F) - q_\alpha^-(F)) + q_\alpha^+(F) \\ &= cq_\alpha^-(F) + (1-c)q_\alpha^+(F),\end{aligned}$$

which completes the proof for (10).

We then show that ρ has convex level sets with respect to $\mathcal{P}$. Suppose that $\rho(F_1) = \rho(F_2)$. Then

$$cq_\alpha^-(F_1) + (1-c)q_\alpha^+(F_1) = cq_\alpha^-(F_2) + (1-c)q_\alpha^+(F_2). \tag{11}$$

For $\lambda \in (0,1)$, define $F_\lambda := \lambda F_1 + (1-\lambda)F_2$. There are two cases:

(i) $c = 0$. Then, $\rho = q_\alpha^+$ and $q_\alpha^+(F_1) = q_\alpha^+(F_2)$. Denote $t = q_\alpha^+(F_1)$, then $F_i(x) > \alpha$ for $x > t$ and $F_i(x) \leq \alpha$ for $x < t$, $i = 1, 2$. Hence, $F_\lambda(x) > \alpha$ for $x > t$ and $F_\lambda(x) \leq \alpha$ for $x < t$, which implies $t = q_\alpha^+(F_\lambda)$, i.e., q_α^+ has convex level sets with respect to $\mathcal{P}$.

(ii) $c \in (0,1)$. Without loss of generality, assume that $q_\alpha^-(F_1) \geq q_\alpha^-(F_2)$. Then it follows from (11) that $q_\alpha^+(F_1) \leq q_\alpha^+(F_2)$. Therefore, $[q_\alpha^-(F_1), q_\alpha^+(F_1)] \subset [q_\alpha^-(F_2), q_\alpha^+(F_2)]$. There are two subcases: (ii.i) $q_\alpha^-(F_1) < q_\alpha^+(F_1)$. In this case, $F_\lambda(x) < \alpha$ for $x <$

$q_\alpha^-(F_1)$; $F_\lambda(x) = \alpha$ for $x \in [q_\alpha^-(F_1), q_\alpha^+(F_1))$; and $F_\lambda(x) > \alpha$ for $x > q_\alpha^+(F_1)$. Therefore, $q_\alpha^-(F_\lambda) = q_\alpha^-(F_1)$ and $q_\alpha^+(F_\lambda) = q_\alpha^+(F_1)$, which implies that $\rho(F_\lambda) = \rho(F_1)$. (ii.ii) $q_\alpha^-(F_1) = q_\alpha^+(F_1)$. In this case, $F_\lambda(x) < \alpha$ for $x < q_\alpha^-(F_1)$ and $F_\lambda(x) > \alpha$ for $x > q_\alpha^+(F_1)$. Therefore, $q_\alpha^-(F_\lambda) = q_\alpha^-(F_1)$ and $q_\alpha^+(F_\lambda) = q_\alpha^+(F_1)$, which implies that $\rho(F_\lambda) = \rho(F_1)$. Therefore, ρ has convex level sets. □

Next, we prove the following Theorem A.1, which shows that among the class of risk measures based on Choquet expected utility theory, only three kinds of risk measures satisfy the necessary condition of being elicitable.

Theorem A.1. *Let $\mathcal{P}_0$ be the set of distributions with finite support. Let h be a distortion function defined on $[0, 1]$ and let $\rho(\cdot)$ be defined as in (2). Then, $\rho(\cdot)$ has convex level sets with respect to $\mathcal{P}_0$ if and only if one of the following three cases holds:*

(i) There exists $\alpha \in [0, 1]$ such that $\rho(F) = VaR_\alpha(F)$, $\forall F$.

(ii) There exists $\alpha \in (0, 1)$ and $c \in [0, 1)$ such that

$$\rho(F) = cq_\alpha^-(F) + (1 - c)q_\alpha^+(F), \ \forall F, \tag{12}$$

where $q_\alpha^-(F) := \inf\{x \mid F(x) \geq \alpha\}$ and $q_\alpha^+(F) := \inf\{x \mid F(x) > \alpha\}$.

(iii) $\rho(F) = \int x dF(x)$, $\forall F$.

Furthermore, the risk measures listed above have convex level sets with respect to $\mathcal{P}$ defined in Theorem 2.1.

Proof. Define $g(u) := 1 - h(1 - u)$, $u \in [0, 1]$. Then $g(0) = 0$, $g(1) = 1$, and g is increasing on $[0, 1]$. And then, ρ can be represented as

$$\rho(F) = -\int_{-\infty}^{0} g(F(x))dx + \int_{0}^{\infty} (1 - g(F(x)))dx.$$

For a discrete distribution $F = \sum_{i=1}^n p_i \delta_{x_i}$, where $0 \leq x_1 < x_2 < \cdots < x_n$, $p_i > 0$, $i = 1, \ldots, n$, and $\sum_{i=1}^n p_i = 1$, it can be shown by simple calculation that $\rho(F) = g(p_1)x_1 + \sum_{i=2}^n (g(\sum_{j=1}^i p_j) - g(\sum_{j=1}^{i-1} p_j))x_i$.

There are three cases for g:

Case (i): for any $q \in (0, 1)$, $g(q) = 0$. In this case, $\rho(F) = \mathrm{VaR}_1(F)$. For any $F_1 \in \mathcal{P}$ and $F_2 \in \mathcal{P}$, if $t = \mathrm{VaR}_1(F_1) = \mathrm{VaR}_1(F_2)$, then $F_i(x) < 1$ for $x < t$ and $F_i(x) = 1$ for $x \geq t$. Hence, for any $F_\lambda := \lambda F_1 + (1 - \lambda)F_2$, $\lambda \in [0, 1]$, it holds that

$F_\lambda(x) < 1$ for $x < t$ and $F_\lambda(x) = 1$ for $x \geq t$. Hence, $\rho(F_\lambda) = \mathrm{VaR}_1(F_\lambda) = t$, i.e., ρ has convex level sets with respect to $\mathcal{P}$.

Case (ii): there exists $q_0 \in (0,1)$ such that $g(q_0) = 1$ and $g(q) \in \{0,1\}$ for all $q \in (0,1)$. Let $\alpha = \inf\{q \mid g(q) = 1\}$. There are three subcases: (ii.i) $\alpha = 0$. Then, $g(u) = 1_{\{u>0\}}$ and $\rho(F) = \mathrm{VaR}_0(F) = \inf\{x \mid F(x) > 0\}$. By similar argument to case (i), we can show that VaR_0 has convex level sets with respect to $\mathcal{P}$; (ii.ii) $\alpha \in (0,1)$ and $g(\alpha) = 1$. Then, $g(u) = 1_{\{u\geq\alpha\}}$ and $\rho(F) = \mathrm{VaR}_\alpha(F)$. For any F_1 and F_2, define $F_\lambda := \lambda F_1 + (1-\lambda)F_2$, $\lambda \in [0,1]$. If $t = F_1^{-1}(\alpha) = F_2^{-1}(\alpha)$, then $F_i(x) < \alpha$ for $x < t$ and $F_i(x) \geq \alpha$ for $x \geq t$, $i = 1,2$. Hence, $F_\lambda(x) < \alpha$ for $x < t$ and $F_\lambda(x) \geq \alpha$ for $x \geq t$, which implies that $\mathrm{VaR}_\alpha(F_\lambda) = t$. Hence, VaR_α has convex level sets with respect to $\mathcal{P}$. (ii.iii) $\alpha \in (0,1)$ and $g(\alpha) = 0$. Then, $g(u) = 1_{\{u>\alpha\}}$ and $\rho(F) = q_\alpha^+(F) := \inf\{x \mid F(x) > \alpha\}$. By Lemma A.2 (with $c = 0$), ρ has convex level sets with respect to $\mathcal{P}$.

Case (iii): there exists $q \in (0,1)$ such that $g(q) \in (0,1)$. Suppose ρ has convex level sets with respect to $\mathcal{P}_0$. For any $0 < x_1 < x_2$ and $q \in (0,1)$ that satisfy

$$1 = \rho(\delta_1) = \rho(q\delta_{x_1} + (1-q)\delta_{x_2}) = x_1 g(q) + x_2(1-g(q)), \tag{13}$$

since ρ has convex level sets, it follows that

$$1 = \rho(v(q\delta_{x_1} + (1-q)\delta_{x_2}) + (1-v)\delta_1), \ \forall v \in (0,1). \tag{14}$$

For any $q \in (0,1)$ such that $g(q) \in (0,1)$, (13) holds for any $(x_1, x_2) = (1-c, -\frac{g(q)}{1-g(q)}(1-c) + \frac{1}{1-g(q)})$, where $c \in (0,1)$. Noting that $x_1 < 1 < x_2$, (14) implies

$$\begin{aligned}
1 &= \rho(v(q\delta_{x_1} + (1-q)\delta_{x_2}) + (1-v)\delta_1) \\
&= x_1 g(vq) + g(vq+1-v) - g(vq) + x_2(1-g(vq+1-v)) \\
&= (1-c)g(vq) + g(vq+1-v) - g(vq) \\
&\quad + \left[-\frac{g(q)}{1-g(q)}(1-c) + \frac{1}{1-g(q)}\right](1-g(vq+1-v)) \\
&= 1 + c\left[-g(vq) + \frac{g(q)}{1-g(q)}(1-g(vq+1-v))\right], \ \forall v \in (0,1), \forall c \in (0,1).
\end{aligned}$$

Therefore,

$$-g(vq) + \frac{g(q)}{1-g(q)}(1-g(vq+1-v)) = 0, \ \forall v \in (0,1), \forall q \text{ such that } g(q) \in (0,1). \tag{15}$$

Let $\alpha = \sup\{q \mid g(q) = 0, q \in [0,1]\}$ and $\beta = \inf\{q \mid g(q) = 1, q \in [0,1]\}$. Since there exists $q_0 \in (0,1)$ such that $g(q_0) \in (0,1)$, it follows that $\alpha \le q_0 < 1$, $g(\alpha) \le g(q_0) < 1$, $\beta \ge q_0 > 0$, and $g(\beta) \ge g(q_0) > 0$.

There are four subcases:

Case (iii.i) $\alpha < \beta$, $g(\alpha) = 0$, and $g(\beta) \in (0,1)$. Since $g(\beta) \in (0,1)$, it follows that $\beta \in (0,1)$. By the definition of β, for any $\eta \in (0, 1-\beta)$, $g(\beta+\eta) = 1$. By the definition of α, $g((\beta+\alpha)/2) > 0$. Hence, $g(\beta-) \ge g((\beta+\alpha)/2) > 0$. Hence, there exists $\epsilon_0 > 0$ such that $g(\beta-\epsilon) > 0$ for any $\epsilon \in (0,\epsilon_0)$. On the other hand, $g(\beta-\epsilon) \le g(\beta) < 1$ for any $\epsilon \in (0,\epsilon_0)$. Hence, $g(\beta-\epsilon) \in (0,1)$ for any $\epsilon \in (0,\epsilon_0)$. Then, for any $\eta \in (0,1-\beta)$ and $\epsilon \in (0,\epsilon_0)$, let $q = \beta - \epsilon$ and $v = \frac{1-\beta-\eta}{1-\beta+\epsilon}$. Then, we have $g(vq+1-v) = g(\beta+\eta) = 1$. Since $g(\beta-\epsilon) \in (0,1)$ for $\epsilon \in (0,\epsilon_0)$, it follows from (15) that $0 = g(vq) = g(\frac{1-\beta-\eta}{1-\beta+\epsilon}(\beta-\epsilon))$, which implies that $g(\beta-) = \lim_{\eta\downarrow 0,\epsilon\downarrow 0} g(\frac{1-\beta-\eta}{1-\beta+\epsilon}(\beta-\epsilon)) = 0$. This contradicts to that $g(\beta-) > 0$. Therefore, this case does not hold.

Case (iii.ii) $\alpha < \beta$ and $g(\alpha) \in (0,1)$. In this case, $\alpha \in (0,1)$. It follows from the definition of β that $g((\alpha+\beta)/2) < 1$. Let $\epsilon_0 = \beta - \alpha$. By the definition of β, $g(\alpha+\epsilon) < 1$ for all $\epsilon \in (0,\epsilon_0)$. In addition, $g(\alpha+\epsilon) \ge g(\alpha) > 0$ for all $\epsilon \in (0,\epsilon_0)$. Hence, $g(\alpha+\epsilon) \in (0,1)$ for all $\epsilon \in (0,\epsilon_0)$. For any $\eta \in (0,\alpha)$ and $\epsilon \in (0,\epsilon_0)$, let $q = \alpha+\epsilon$ and $v = \frac{\alpha-\eta}{\alpha+\epsilon}$. Then it follows from the definition of α that $g(vq) = g(\alpha-\eta) = 0$, which implies from (15) that $1 = g(vq+1-v) = g(\alpha-\eta+\frac{\epsilon+\eta}{\alpha+\epsilon})$, for any $\epsilon \in (0,\epsilon_0), \eta \in (0,\alpha)$. Then, $g(\alpha+) = \lim_{\epsilon\downarrow 0,\eta\downarrow 0} g(\alpha-\eta+\frac{\epsilon+\eta}{\alpha+\epsilon}) = 1$, which contradicts to $g(\alpha+) \le g((\alpha+\beta)/2) < 1$. Therefore, this case does not hold.

Case (iii.iii) $\alpha = \beta$ and $g(\alpha) = c \in (0,1)$. In this case, $\alpha = \beta \in (0,1)$. By the definition of α and β, $g(x) = 0$ for $x < \alpha$ and $g(x) = 1$ for $x > \alpha$. By Lemma A.2, ρ has convex level sets with respect to $\mathcal{P}$.

Case (iii.iv) $\alpha < \beta$, $g(\alpha) = 0$, $g(\beta) = 1$, and there exists q_0 such that $g(q_0) \in (0,1)$. In this case, $\alpha < q_0 < \beta$. We will show that that $g(u) = u$ for all $u \in (0,1)$.

First, we will show that $\alpha = 0$ and $\beta = 1$. Suppose for the sake of contradiction that $\alpha > 0$. Since $\alpha < q_0$, it follows that $g(\alpha+\epsilon) < 1$ for all $\epsilon \in (0,\epsilon_0)$, where $\epsilon_0 = q_0 - \alpha$. Furthermore, by the definition of α, $g(\alpha+\epsilon) > 0$ for all $\epsilon \in (0,\epsilon_0)$. Hence, $g(\alpha+\epsilon) \in (0,1)$ for all $\epsilon \in (0,\epsilon_0)$. For any $\eta \in (0,\alpha)$ and $\epsilon \in (0,\epsilon_0)$, let $q = \alpha+\epsilon$ and $v = \frac{\alpha-\eta}{\alpha+\epsilon}$. Then it follows from the definition of α that $g(vq) = g(\alpha-\eta) = 0$, which implies from (15) that $1 = g(vq+1-v) = g(\alpha-\eta+\frac{\epsilon+\eta}{\alpha+\epsilon})$, for any $\epsilon \in (0,\epsilon_0), \eta \in (0,\alpha)$. Then, $g(\alpha+) = \lim_{\epsilon\downarrow 0,\eta\downarrow 0} g(\alpha-\eta+\frac{\epsilon+\eta}{\alpha+\epsilon}) = 1$, which contradicts to $g(\alpha+) \le g(q_0) < 1$.

Therefore, $\alpha = 0$.

In addition, suppose for the sake of contradiction that $\beta < 1$. Then, by the definition of β, for any $\eta \in (0, 1-\beta)$, $g(\beta+\eta) = 1$. Let $\epsilon_0 = \beta - q_0$. Since $\beta > q_0$, $g(\beta-\epsilon) \geq g(q_0) > 0$ for any $\epsilon \in (0, \epsilon_0)$. By the definition of β, $g(\beta-\epsilon) < 1$ for any $\epsilon \in (0, \epsilon_0)$. Hence, $g(\beta-\epsilon) \in (0,1)$ for any $\epsilon \in (0, \epsilon_0)$. Then, for any $\eta \in (0, 1-\beta)$ and $\epsilon \in (0, \epsilon_0)$, let $q = \beta - \epsilon$ and $v = \frac{1-\beta-\eta}{1-\beta+\epsilon}$. Then, we have $g(vq+1-v) = g(\beta+\eta) = 1$. Since $g(\beta-\epsilon) \in (0,1)$ for any $\epsilon \in (0, \epsilon_0)$, it follows from (15) that $0 = g(vq) = g(\frac{1-\beta-\eta}{1-\beta+\epsilon}(\beta-\epsilon))$, which implies that $g(\beta-) = \lim_{\eta\downarrow 0, \epsilon\downarrow 0} g(\frac{1-\beta-\eta}{1-\beta+\epsilon}(\beta-\epsilon)) = 0$. This contradicts to that $g(\beta-) \geq g(q_0) > 0$. Therefore, $\beta = 1$.

Then, it follows from $\alpha = 0$ and $\beta = 1$ that

$$g(q) \in (0,1),\ \forall q \in (0,1). \tag{16}$$

Therefore, it follows from (15) and (16) that

$$-g(vq) + \frac{g(q)}{1-g(q)}(1 - g(vq+1-v)) = 0,\ \forall v \in (0,1), \forall q \in (0,1). \tag{17}$$

Second, for any $q \in (0,1)$ and $v \in (0,1)$, $vq+1-v > q$ and $\lim_{v\uparrow 1}(vq+1-v) = q$. Then, it follows from (17) that

$$\begin{aligned} g(q-) &= \lim_{v\uparrow 1} g(vq) = \lim_{v\uparrow 1} \frac{g(q)}{1-g(q)}(1 - g(vq+1-v)) \\ &= \frac{g(q)}{1-g(q)}(1 - g(q+)),\ \forall q \in (0,1). \end{aligned} \tag{18}$$

Third, we will show that g is strictly increasing on $(0,1)$. Suppose for the sake of contradiction that there exist $0 < u_1 < u_2 < 1$ such that $g(u_1) = g(u_2)$. Since $u_1 \in (0,1)$, it follows from (16) that $g(u_1) > 0$. Let $u_0 = \inf\{u \mid g(u) = g(u_1)\}$. Then since $g(u_0) = g(u_1) > 0$, $u_0 > 0$. Since $\lim_{v\uparrow 1} vu_0 + 1 - v = u_0$, there exists $v \in (0,1)$ such that $vu_0 + 1 - v \in (u_0, u_2)$. Hence,

$$\frac{g(u_0)}{1-g(u_0)}(1 - g(vu_0+1-v)) = g(u_0) > g(vu_0),$$

which contradicts to (17). Therefore, g is strictly increasing on $(0,1)$, and then $g(p_1) - g(p_2) \neq 0$ for any $p_1 \neq p_2$.

Fourth, we will show that $g(1-) = 1$ and $g(0+) = 0$. Consider $0 < x_1 < x_2 < x_3$ and $p_1, p_2 \in (0,1)$ such that

$$\rho(p_1\delta_{x_1} + (1-p_1)\delta_{x_2}) = \rho(p_2\delta_{x_1} + (1-p_2)\delta_{x_3}),$$

which is equivalent to

$$x_1g(p_1) + x_2(1 - g(p_1)) = x_1g(p_2) + (1 - g(p_2))x_3. \tag{19}$$

Let $\frac{x_1}{x_2} = c_1$ and $\frac{x_3}{x_2} = c_3$. Then, $c_1 \in (0, 1)$, $c_3 > 1$, and (19) is equivalent to

$$c_1 = \frac{1 - g(p_2)}{g(p_1) - g(p_2)}c_3 - \frac{1 - g(p_1)}{g(p_1) - g(p_2)}. \tag{20}$$

For any fixed $0 < p_1 < p_2 < 1$ and $1 < c_3 < \frac{1-g(p_1)}{1-g(p_2)}$, define c_1 as in (20). Then, $c_1 \in (0, 1)$. For any such p_1, p_2, c_3, and c_1, it follows from the convexity of the level sets of ρ that

$$\begin{aligned}
&x_1g(p_1) + x_2(1 - g(p_1)) = \rho(p_1\delta_{x_1} + (1 - p_1)\delta_{x_2})\\
&= \rho(v(p_1\delta_{x_1} + (1 - p_1)\delta_{x_2}) + (1 - v)(p_2\delta_{x_1} + (1 - p_2)\delta_{x_3}))\\
&= \rho((vp_1 + (1 - v)p_2)\delta_{x_1} + v(1 - p_1)\delta_{x_2} + (1 - v)(1 - p_2)\delta_{x_3})\\
&= x_1g(vp_1 + (1 - v)p_2) + x_2(g(v + (1 - v)p_2) - g(vp_1 + (1 - v)p_2))\\
&\quad + x_3(1 - g(v + (1 - v)p_2)),\ \forall v \in (0, 1),
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&c_1[g(p_1) - g(vp_1 + (1 - v)p_2)] + 1 - g(p_1) - g(v + (1 - v)p_2) + g(vp_1 + (1 - v)p_2)\\
&= c_3[1 - g(v + (1 - v)p_2)],\ \forall v \in (0, 1).
\end{aligned}$$

Plugging (20) into the above equation, we obtain that for any $0 < p_1 < p_2 < 1$, any $1 < c_3 < \frac{1-g(p_1)}{1-g(p_2)}$, and any $v \in (0, 1)$, it holds that

$$\begin{aligned}
0 = {}& c_3\left[\frac{1 - g(p_2)}{g(p_1) - g(p_2)}(g(p_1) - g(vp_1 + (1 - v)p_2)) - 1 + g(v + (1 - v)p_2)\right]\\
&- \frac{1 - g(p_1)}{g(p_1) - g(p_2)}[g(p_1) - g(vp_1 + (1 - v)p_2)] + 1 - g(p_1)\\
&- g(v + (1 - v)p_2) + g(vp_1 + (1 - v)p_2).
\end{aligned} \tag{21}$$

Therefore,

$$\begin{aligned}
0 = {}& - \frac{1 - g(p_1)}{g(p_1) - g(p_2)}[g(p_1) - g(vp_1 + (1 - v)p_2)] + 1 - g(p_1)\\
&- g(v + (1 - v)p_2) + g(vp_1 + (1 - v)p_2),\ \forall v \in (0, 1), \forall p_1 < p_2,
\end{aligned}$$

which is equivalent to

$$\begin{aligned}0 = g(vp_1 + (1-v)p_2)(1-g(p_2)) + g(v+(1-v)p_2)(g(p_2)-g(p_1)) \\ + g(p_1)g(p_2) - g(p_2),\ \forall v \in (0,1), \forall p_1 < p_2. \qquad (22)\end{aligned}$$

Letting $v \uparrow 1$ in (22), we obtain

$$0 = g(p_1+)(1-g(p_2)) + g(1-)(g(p_2)-g(p_1)) + g(p_1)g(p_2) - g(p_2), \forall p_1 < p_2. \quad (23)$$

Since g is increasing on $(0,1)$, there exists $p_1^* \in (0,1)$, such that g is continuous at p_1^*. Choose any $p_2^* > p_1^*$. Letting $p_1 = p_1^*$ and $p_2 = p_2^*$ in (23) leads to $(g(p_1^*) - g(p_2^*))(1 - g(1-)) = 0$. Since g is strictly increasing, it follows that

$$g(1-) = 1. \qquad (24)$$

Letting $q = \frac{1}{2}$ in (17) leads to

$$\frac{g(\frac{v}{2})}{1-g(1-\frac{v}{2})} = \frac{g(\frac{1}{2})}{1-g(\frac{1}{2})}, \forall v \in (0,1). \qquad (25)$$

It follows from (25) and (24) that

$$\begin{aligned}g(0+) &= \lim_{v\downarrow 0} g(\frac{v}{2}) = \lim_{v\downarrow 0} \frac{g(\frac{1}{2})}{1-g(\frac{1}{2})}(1-g(1-\frac{v}{2})) \\ &= \frac{g(\frac{1}{2})}{1-g(\frac{1}{2})}(1-g(1-)) \\ &= 0. \qquad (26)\end{aligned}$$

Fifth, we will show that g is continuous on $(0,1)$. By (17), we have

$$\begin{aligned}g(v-) &= \lim_{q\uparrow 1} g(vq) = \lim_{q\uparrow 1} \frac{g(q)}{1-g(q)}(1-g(vq+1-v)) \\ &= \lim_{q\uparrow 1} g(q) \lim_{q\uparrow 1} \frac{1-g(vq+1-v)}{1-g(q)} \\ &= g(1-) \lim_{q\uparrow 1} \frac{1-g(vq+1-v)}{g((1-q)v)} \frac{g((1-q)v)}{g(1-q)} \frac{g(1-q)}{1-g(q)} \\ &= \lim_{q\uparrow 1} \frac{1-g(\frac{1}{2})}{g(\frac{1}{2})} \frac{g((1-q)v)}{g(1-q)} \frac{g(\frac{1}{2})}{1-g(\frac{1}{2})} \text{ (by (24) and (25))} \\ &= \lim_{q\uparrow 1} \frac{g((1-q)v)}{g(1-q)} \\ &= \lim_{q\downarrow 0} \frac{g(qv)}{g(q)},\ \forall v \in (0,1). \qquad (27)\end{aligned}$$

Now consider $0 = x_1 < x_2 < x_3 < x_4$ and $p_1, p_2 \in (0, 1)$ such that

$$\rho(p_1\delta_{x_1} + (1 - p_1)\delta_{x_3}) = \rho(p_2\delta_{x_2} + (1 - p_2)\delta_{x_4}),$$

which is equivalent to

$$x_1 g(p_1) + x_3(1 - g(p_1)) = x_2 g(p_2) + x_4(1 - g(p_2)). \tag{28}$$

Since ρ has convex level sets, it follows that for any $v \in (0, 1)$, it holds that

$$\begin{aligned}
&x_3(1 - g(p_1)) = x_1 g(p_1) + x_3(1 - g(p_1)) = \rho(p_1\delta_{x_1} + (1 - p_1)\delta_{x_3}) \\
&= \rho(v(p_1\delta_{x_1} + (1 - p_1)\delta_{x_3}) + (1 - v)(p_2\delta_{x_2} + (1 - p_2)\delta_{x_4})) \\
&= \rho(vp_1\delta_{x_1} + (1 - v)p_2\delta_{x_2} + v(1 - p_1)\delta_{x_3} + (1 - v)(1 - p_2)\delta_{x_4}) \\
&= x_2(g(vp_1 + (1 - v)p_2) - g(vp_1)) \\
&\quad + x_3(g(v + (1 - v)p_2) - g(vp_1 + (1 - v)p_2)) + x_4(1 - g(v + (1 - v)p_2)).
\end{aligned} \tag{29}$$

Let $\frac{x_3}{x_2} = 1 + c_3$ and $\frac{x_4}{x_2} = 1 + c_3 + c_4$. Then, $c_3 > 0$, $c_4 > 0$, and (28) becomes

$$c_3 = \frac{1 - g(p_2)}{g(p_2) - g(p_1)} c_4 + \frac{g(p_1)}{g(p_2) - g(p_1)}. \tag{30}$$

Furthermore, (29) is equivalent to

$$\begin{aligned}
0 &= g(vp_1 + (1 - v)p_2) - g(vp_1) + (1 + c_3 + c_4)(1 - g(v + (1 - v)p_2)) \\
&\quad + (1 + c_3)(g(v + (1 - v)p_2) - g(vp_1 + (1 - v)p_2) - 1 + g(p_1)), \forall v \in (0, 1).
\end{aligned} \tag{31}$$

For any $0 < p_1 < p_2 < 1$ and $c_4 > 0$, let c_3 be defined in (30). Then, $c_3 > 0$. Hence, (31) holds for any such p_1, p_2, c_3, and c_4. Plugging (30) into (31), we obtain that for any $0 < p_1 < p_2 < 1$ and any $c_4 > 0$, it holds that

$$\begin{aligned}
0 &= g(vp_1 + (1 - v)p_2) - g(vp_1) + \frac{g(p_2)}{g(p_2) - g(p_1)}[g(p_1) - g(vp_1 + (1 - v)p_2)] \\
&\quad + c_4 \frac{1 - g(p_2)}{g(p_2) - g(p_1)}[g(v + (1 - v)p_2) - g(vp_1 + (1 - v)p_2) - 1 + g(p_1)] \\
&\quad + c_4 \frac{1 - g(p_1)}{g(p_2) - g(p_1)}[1 - g(v + (1 - v)p_2)], \forall v \in (0, 1),
\end{aligned} \tag{32}$$

which implies that

$$\begin{aligned}
0 &= g(vp_1 + (1 - v)p_2) - g(vp_1) \\
&\quad + \frac{g(p_2)}{g(p_2) - g(p_1)}[g(p_1) - g(vp_1 + (1 - v)p_2)], \ \forall 0 < p_1 < p_2 < 1, \forall v \in (0, 1),
\end{aligned}$$

which can be simplified to be

$$-g(vp_1+(1-v)p_2)-(g(p_2)-g(p_1))\frac{g(vp_1)}{g(p_1)}+g(p_2)=0,\ \forall p_1<p_2,\forall v\in(0,1). \quad (33)$$

Letting $p_2\uparrow 1$ in (33) and applying (24), we obtain

$$-g((vp_1+1-v)-)-(1-g(p_1))\frac{g(vp_1)}{g(p_1)}+1=0,\ \forall 0<p_1<1,\forall v\in(0,1). \quad (34)$$

Then, it follows from (17) and (34) that

$$g((vp_1+1-v)-)=g(vp_1+1-v),\forall 0<p_1<1,\forall v\in(0,1),$$

which implies that

$$g(v-)=g(v),\forall v\in(0,1). \quad (35)$$

It follows from (18) and (35) that g is continuous on $(0,1)$, i.e.,

$$g(v-)=g(v)=g(v+),\forall v\in(0,1). \quad (36)$$

Lastly, we will show that $g(u)=u$ for any $u\in(0,1)$. Letting $p_1\downarrow 0$ in (33), we obtain

$$-g(((1-v)p_2)+)-(g(p_2)-g(0+))\lim_{p_1\downarrow 0}\frac{g(vp_1)}{g(p_1)}+g(p_2)=0,\ \forall 0<p_2<1,\forall v\in(0,1). \quad (37)$$

Applying (26), (27), and (36) to (37), we obtain

$$g((1-v)p_2)=g(p_2)(1-g(v)),\ \forall 0<p_2<1,\forall v\in(0,1). \quad (38)$$

Letting $p_2\uparrow 1$ in (38) and using (24) and (36), we obtain

$$g(1-v)=g(1-)(1-g(v))=1-g(v),\ \forall v\in(0,1), \quad (39)$$

which in combination with (38) implies

$$g(vp_2)=g(v)g(p_2),\ \forall 0<p_2<1,\forall v\in(0,1). \quad (40)$$

In the following, we will show by induction that

$$g(\frac{k}{2^n})=\frac{k}{2^n},\ k=1,2,\ldots,2^n-1,\forall n\in\mathbb{N}. \quad (41)$$

Letting $v = \frac{1}{2}$ in (39), we obtain $g(\frac{1}{2}) = \frac{1}{2}$. Hence, (41) holds for $n = 1$. Suppose (41) holds for n. We will show that it also holds for $n+1$. In fact, for any $0 \le k \le 2^{n-1}-1$, since $1 \le 2k+1 \le 2^n - 1$, it follows from (40) that

$$g(\frac{2k+1}{2^{n+1}}) = g(\frac{1}{2})g(\frac{2k+1}{2^n}) = \frac{2k+1}{2^{n+1}},\ 0 \le k \le 2^{n-1} - 1. \tag{42}$$

For any $2^{n-1} \le k \le 2^n - 1$, it holds that $1 \le 2^{n+1} - (2k+1) \le 2^n - 1$. Hence, it follows from (39) that

$$\begin{aligned} & g(\frac{2k+1}{2^{n+1}}) = 1 - g(\frac{2^{n+1} - (2k+1)}{2^{n+1}}) = 1 - \frac{2^{n+1} - (2k+1)}{2^{n+1}} \text{ (by (42))} \\ = & \frac{2k+1}{2^{n+1}},\ 2^{n-1} \le k \le 2^n - 1. \end{aligned} \tag{43}$$

In addition, for any $1 \le k \le 2^n - 1$, $g(\frac{2k}{2^{n+1}}) = g(\frac{k}{2^n}) = \frac{k}{2^n}$, which in combination with (42) and (43) implies that (41) holds for $n+1$, and hence holds for any n. Since $\{k/2^n, k = 1, \ldots, 2^n - 1, n \in \mathbb{N}\}$ is dense on $(0,1)$ and g is continuous on $(0,1)$, it follows from (41) that $g(u) = u$ for all $u \in (0,1)$, which completes the proof. □

Finally, the proof of Theorem 2.1 is as follows.

Proof of Theorem 2.1. By Lemma A.1 and Theorem A.1, only those risk measures listed in cases (i)-(iii) of Theorem A.1 satisfy the necessary condition for being an elicitable risk measure with respect to $\mathcal{P}$. Therefore, we only need to study the elicitability of those risk measures.

First, we will show that $\rho(F) = \mathrm{VaR}_0(F) = \inf\{x \mid F(x) > 0\}$ is not elicitable. Suppose for the sake of contradiction that VaR_0 is elicitable, then there exists a function S such that (5) holds. For any u, letting $F = \delta_u$ in (5) yields

$$S(u,u) \le S(x,u), \forall x, \forall u, \text{ and the equality holds only if } u \le x. \tag{44}$$

For any $u < v$ and $p \in (0,1)$, letting $F = p\delta_u + (1-p)\delta_v$ in (5) yields $pS(u,u) + (1-p)S(u,v) \le pS(x,u) + (1-p)S(x,v)$, $\forall x$. Letting $p \to 0$ leads to

$$S(u,v) \le S(x,v), \forall u < v, \forall x. \tag{45}$$

Letting $x = v$ in (45), we obtain

$$S(u,v) \le S(v,v), \forall u < v. \tag{46}$$

By (44), $S(v,v) \le S(u,v)$, $\forall u$, which in combination with (46) implies $S(u,v) = S(v,v)$, $\forall u < v$; however, by (44), $S(u,v) = S(v,v)$ implies $v \le u$; this is a contradiction. Hence, VaR_0 is not elicitable.

Second, we will show that for any $\alpha \in (0,1)$, $\mathrm{VaR}_\alpha(F)$ is elicitable with respect to $\mathcal{P}$. Define

$$S(x,y) = (1_{\{x \ge y\}} - \alpha)(x-y). \tag{47}$$

It follows from Theorem 9 in Gneiting (2011) that $S(x,y)$ is a strictly consistent forecasting objective function for the set-valued α-quantile functional. More precisely,

$$[q_\alpha^-(F), q_\alpha^+(F)] = \arg\min_x \int S(x,y)dF(y),$$

where $q_\alpha^-(F) := \inf\{y \mid F(y) \ge \alpha\}$ and $q_\alpha^+(F) := \inf\{y \mid F(y) > \alpha\}$. Therefore, $\mathrm{VaR}_\alpha(F) = q_\alpha^-(F)$ satisfies (5) with S defined in (47).

Third, we will show that $\rho(F) = \mathrm{VaR}_1(F)$ is elicitable with respect to $\mathcal{P}$. Let $a > 0$ be a constant and define the forecasting objective function

$$S(x,y) = \begin{cases} 0, & \text{if } x \ge y; \\ a, & \text{else.} \end{cases}$$

Then for any $F \in \mathcal{P}$ and any $x \ge \rho(F)$,

$$\int_{\mathbb{R}} S(x,y)dF(y) = \int_{y \le \rho(F)} S(x,y)dF(y) = 0.$$

On the other hand, for any $F \in \mathcal{P}$ and any $x < \rho(F)$,

$$\int_{\mathbb{R}} S(x,y)dF(y) = \int_{x<y\le\rho(F)} S(x,y)dF(y) = a\int_{x<y\le\rho(F)} dF(y) > 0.$$

Therefore, for any $F \in \mathcal{P}$, $\rho(F) = \min\{x \mid x \in \arg\min_x \int S(x,y)dF(y)\}$.

Fourth, we will show that ρ defined in (12) is not elicitable with respect to $\mathcal{P}$. Suppose for the purpose of contradiction that ρ is elicitable. Fix any $a > 0$ and denote $I := (-a, a)$. Let $\mathcal{P}_I$ be the set of probability measures that have strictly positive probability density on the interval I and whose support is I. Then since $\mathcal{P}_I \subset \mathcal{P}$ and ρ is elicitable with respect to $\mathcal{P}$, ρ is also elicitable with respect to $\mathcal{P}_I$. Therefore, there exists a forecasting objective function $S(x,y)$ such that

$$\rho(F) = \min\{x \mid x \in \arg\min_x \int S(x,y)dF(y)\}, \forall F \in \mathcal{P}_I.$$

For any $F \in \mathcal{P}_I$, the equation $F(x) = \alpha$ has a unique solution $q_\alpha(F)$ and $q_\alpha^-(F) = q_\alpha(F) = q_\alpha^+(F)$. Hence, $\rho(F) = q_\alpha(F), \forall F \in \mathcal{P}_I$. Therefore, we have

$$q_\alpha(F) \in \arg\min_x \int S(x,y)dF(y), \forall F \in \mathcal{P}_I.$$

Then, it follows from the proposition in Thomson (1979, p. 372) that[20] there exists measurable functions A_1, A_2, B_1, and B_2 such that

$$S(x,y) = \begin{cases} A_1(x) + B_1(y) \text{ } a.e. \text{ if } y \le x, \\ A_2(x) + B_2(y) \text{ } a.e. \text{ if } y > x, \end{cases} \tag{48}$$

and

$$(A_1(x_1) - A_1(x_2))\alpha + (A_2(x_1) - A_2(x_2))(1-\alpha) = 0, \ \forall x_1, x_2 \in I. \tag{49}$$

Choose a distribution $F_0 \in \mathcal{P}$ such that $q_\alpha^-(F_0) < q_\alpha^+(F_0)$, F_0 has a density f_0 that satisfies $f_0(x) = 0$ for $x \in [q_\alpha^-(F_0), q_\alpha^+(F_0)]$, and $F_0(q_\alpha^-(F_0)) = F_0(q_\alpha^+(F_0)) = \alpha$. Then, it follows from (48) that

$$\begin{aligned}
&\int S(x,y)dF_0(y) = \int_{y \le x} S(x,y)f_0(y)dy + \int_{y>x} S(x,y)f_0(y)dy \\
&= \int_{y \le x} (A_1(x) + B_1(y))f_0(y)dy + \int_{y>x} (A_2(x) + B_2(y))f_0(y)dy \\
&= A_1(x)\int_{y \le x} f_0(y)dy + \int_{y \le x} B_1(y)f_0(y)dy + A_2(x)\int_{y>x} f_0(y)dy + \int_{y>x} B_2(y)f_0(y)dy \\
&= A_1(x)\alpha + \int_{y \le x} B_1(y)f_0(y)dy + A_2(x)(1-\alpha) + \int_{y>x} B_2(y)f_0(y)dy.
\end{aligned} \tag{50}$$

Since $f_0(x) = 0$ for $x \in [q_\alpha^-(F_0), q_\alpha^+(F_0)]$, it follows that

$$\int_{y \le x_1} B_1(y)f_0(y)dy = \int_{y \le x_2} B_1(y)f_0(y)dy, \ \forall x_1, x_2 \in [q_\alpha^-(F_0), q_\alpha^+(F_0)], \tag{51}$$

$$\int_{y > x_1} B_2(y)f_0(y)dy = \int_{y > x_2} B_2(y)f_0(y)dy, \ \forall x_1, x_2 \in [q_\alpha^-(F_0), q_\alpha^+(F_0)]. \tag{52}$$

Then, for any $x \in [q_\alpha^-(F_0), q_\alpha^+(F_0)]$, it follows from (48), (49), (50), and (51) that

$$\int S(x,y)dF_0(y) = \int S(\rho(F_0),y)dF_0(y), \ \forall x \in [q_\alpha^-(F_0), q_\alpha^+(F_0)],$$

[20] Thomson (1979) obtain the proposition for the case when the interval $I = (-\infty, \infty)$; in our case, $I = (-a, a)$. It can be verified that the proof for the proposition in Thomson (1979) can be easily adapted to the case of $I = (-a, a)$. The details are available from the authors upon request.

which implies that

$$[q_\alpha^-(F_0), q_\alpha^+(F_0)] \subset \arg\min_x \int S(x,y)dF_0(y).$$

This contradicts to that

$$\rho(F_0) = cq_\alpha^-(F_0) + (1-c)q_\alpha^+(F_0) = \min\{x \mid x \in \arg\min_x \int S(x,y)dF_0(y)\}.$$

Therefore, ρ defined in (12) is not elicitable.

Fifth, it follows from Theorem 7 in Gneiting (2011) that $\rho(F) := \int x dF(x)$ is elicitable with respect to $\mathcal{P}$. The proof is thus completed. $\square$

References

Acerbi (2002). Spectral measures of risk: A coherent representation of subjective risk aversion, *Journal of Banking & Finance* **26**(7): 1505–1518.

Adrian, T. and Shin, H. S. (2014). Procyclical leverage and Value-at-Risk, *Review of Financial Studies* **27**(2): 373–403.

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

Aumann, R. J. and Serrano, R. (2008). An economic index of riskiness, *Journal of Political Economy* **116**(5): 810–836.

Basel Committee on Banking Supervision (2009). Revisions to the Basel II market risk framework, *Document*, Bank for International Settlements, Basel, Switzerland.

Basel Committee on Banking Supervision (2013). Fundamental review of the trading book: A revised market risk framework, *Consultative Document*, Bank for International Settlements, Basel, Switzerland.

Bellini, F. and Bignozzi, V. (2013). Elicitable risk measures, *Preprint*, Università di Milano-Bicocca and ETH Zurich.

Brunnermeier, M. K., Crocket, A., Goodhart, C., Perssaud, A. and Shin, H. S. (2009). *The Fundamental Principles of Financial Regulation: 11th Geneva Report on the World Economy*.

Brunnermeier, M. K. and Pedersen, L. H. (2009). Market liquidity and funding liquidity, *Review of Financial Studies* **22**(6): 2201–2238.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and Montrucchio, L. (2013a). Ambiguity and robust statistics, *Journal of Economic Theory* **148**(3): 974 – 1049.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and Montrucchio, L. (2013b). Choquet integration on Riesz spaces and dual comonotonicity, *Transactions of the American Mathematical Society*, forthcoming.

Cherny, A. and Madan, D. (2009). New measures for performance evaluation, *Review of Financial Studies* **22**(7): 2571–2606.

Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures, *Quantitative Finance* **10**(6): 593–606.

Daníelsson, J., Jorgensen, B. N., Samorodnitsky, G., Sarma, M. and de Vries, C. G. (2013). Fat tails, VaR and subadditivity, *Journal of Econometrics* **172**(2): 283–291.

Davis, M. H. A. (2013). Consistency of risk measure estimates, *Preprint*, Imperial College London.

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

Duffie, D. and Pan, J. (1997). An overview of Value at Risk, *Journal of Derivatives* **4**(3): 7–49. Reprinted in *Options Markets*, edited by G. Constantinides and A. G. Malliaris, London: Edward Elgar, 2001.

Duffie, D. and Pan, J. (2001). Analytical Value-at-Risk with jumps and credit risk, *Finance and Stochastics* **5**(2): 115–180.

Embrechts, P. and Hofert, M. (2014). Statistics and quantitative risk management for banking and insurance, *Annual Review of Statistics and Its Application* **1**(1): 493–514.

Engelberg, J., Manski, C. F. and Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters, *Journal of Business & Economic Statistics* **27**(1): 30–41.

Fama, E. T. and Miller, M. H. (1972). *The Theory of Finance*, Dryden Press.

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

Foster, D. P. and Hart, S. (2009). An operational measure of riskiness, *Journal of Political Economy* **117**(5): 785–814.

Foster, D. P. and Hart, S. (2012). A wealth-requirement axiomatization of riskiness, *Preprint*, University of Pennsylvania and the Hebrew University of Jerusalem.

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

Gaglianone, W. P., Lima, L. R., Linton, O. and Smith, D. R. (2011). Evaluating Value-at-Risk models via quantile regression, *Journal of Business & Economic S-tatistics* **29**(1): 150–160.

Gilboa, I., Maccheroni, F., Marinacci, M. and Schmeidler, D. (2010). Objective and subjective rationality in a multiple prior model, *Econometrica* **78**: 755–770.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics* **18**(2): 141–153.

Gneiting, T. (2011). Making and evaluating point forecasts, *Journal of the American Statistical Association* **106**(494): 746–762.

Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules, *Journal of Financial Intermediation* **12**(3): 199–232.

Gordy, M. B. and Howells, B. (2006). Procyclicality in Basel II: Can we treat the disease without killing the patient? *Journal of Financial Intermediation* **15**: 395–417.

Hampel, F. R. (1971). A general qualitative definition of robustness, *The Annuals of Mathematical Statistics* **42**(6): 1887–1896.

Hansen, L. P. (2013). Challenges in identifying and measuring systemic risk, *in* M. Brunnermeier and A. Krishnamurthy (eds), *Risk Topography: Systemic Risk and Macro Modeling*, University of Chicago Press, chapter 1.

Hansen, L. P. and Sargent, T. J. (2001). Robust control and model uncertainty, *American Economic Review* **91**(2): 60–66.

Hansen, L. P. and Sargent, T. J. (2007). *Robustness*, Princeton Unviersity Press.

Huber, P. J. (1981). *Robust Statistics*, 2nd edn, John Wiley & Sons, Hoboken, NJ.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics*, 2nd edn, John Wiley & Sons, Hoboken, NJ.

Ibragimov, R. (2009). Portfolio diversification and value at risk under thick-tailedness, *Quantitative Finance* **9**(5): 565–580.

Ibragimov, R. and Walden, J. (2007). The limits of diversification when losses may be large, *Journal of Banking & Finance* **31**(8): 2551–2569.

Jorion, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk*, 3 edn, McGraw-Hill, Boston.

Kou, S., Peng, X. and Heyde, C. C. (2013). External risk measures and Basel accords, *Mathematics of Operations Research* **38**(3): 393–417.

Linton, O. and Xiao, Z. (2013). Estimation of and inference about the expected shortfall for time series with infinite variance, *Econometric Theory* **29**(4): 771–807.

Maccheroni, F., Marinacci, M. and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences, *Econometrica* **74**(6): 1447–1498.

Moscadelli, M. (2004). The modelling of operational risk: experience with the analysis of the data collected by the basel committee, *Preprint 517*, Banca d'Italia.

Osband, K. H. (1985). *Providing Incentives for Better Cost Forecasting*, PhD thesis, University of California at Berkeley.

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

Savage, L. J. (1971). Elicitation of personal probabilities and expectations, *Journal of the American Statistical Association* **66**(336): 783–810.

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

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

So, M. K. P. and Wong, C.-M. (2012). Estimation of multiple period expected shortfall and median shortfall for risk management, *Quantitative Finance* **12**(5): 739–754.

Song, Y. and Yan, J.-A. (2006). The representation of two types of functionals on $l^\infty(\omega, \mathcal{F})$ and $l^\infty(\omega, \mathcal{F}, \mathbb{P})$, *Science in China Series A: Mathematics* **49**(10): 1376–1382.

Song, Y. and Yan, J.-A. (2009). Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders, *Insurance: Mathematics and Economics* **45**: 459–465.

Tasche, D. (2002). Expected shortfall and beyond, *Journal of Banking & Finance* **26**(7): 1519–1533.

Thomson, W. (1979). Eliciting production possibilities from a well-informed manager, *Journal of Economic Theory* **20**(3): 360–380.

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

Wen, Z., Peng, X., Liu, X., Bai, X. and Sun, X. (2013). Asset allocation under the Basel Accord risk measures, *Preprint*, Peking University.

Xi, J., Coleman, T. F. and Li, Y. (2013). A gradual non-convexification method for minimizing VaR, *Journal of Risk*, forthcoming.

Xia, J. (2013). Comonotonic convex preferences, *Preprint*, Chinese Academy of Sciences.

Ziegel, J. F. (2013). Coherence and elicitability, *Mathematical Finance*, forthcoming.

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 38, No. 3, August 2013, pp. 393–417
ISSN 0364-765X (print) | ISSN 1526-5471 (online)
http://dx.doi.org/10.1287/moor.1120.0577

External Risk Measures and Basel Accords

Steven Kou
Department of Industrial Engineering and Operations Research, 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 Accords, which use value-at-risk with scenario analysis as the risk measures for setting capital requirements. We argue that 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 is 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; expected shortfall; median shortfall; value-at-risk
MSC2000 subject classification: Primary: 91B30, 62P20; Secondary: 91B08
OR/MS subject classification: Primary: regulations, risk; Secondary: banks
History: Received September 8, 2011; revised March 20, 2012. Published online in *Articles in Advance* March 13, 2013.

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 Ω be the set of all the possible states of nature at the end of an observation period, and $\mathscr{X}$ be the set of financial losses, which are random variables defined on Ω. Then a risk measure ρ is a mapping from $\mathscr{X}$ 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. Because 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 et al. [32] 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 predefined 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 α is defined as $\mathrm{VaR}_\alpha(X) := \inf\{x \mid F(x) \geq \alpha\} = F^{-1}(\alpha)$. In practice, $\mathrm{VaR}_\alpha(X)$ is usually estimated from a sample of X, i.e., a data set $\tilde{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Gordy [19] 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 [6, 8] 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 §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 et al. [49], we shall define a new class of risk measures that satisfy a new set of axioms. The new class of risk measures includes 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 countercyclical indexing risk measure suggested by Gordy and Howells [20].

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, whereas 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 §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 (§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 (§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 (§§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 (§§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 (§5). (vi) We show that median shortfall, a special case of natural risk statistics, is more robust than expected shortfall suggested by coherent risk measures (§5.4). (vii) We show that natural risk statistics include a subclass of robust risk measures that are suitable for external regulation (§5.5). (viii) We provide other critiques of the subadditivity axiom of coherent risk measures from the viewpoints of diversification and bankruptcy protection (§6). (ix) We derive the Euler capital allocation rule under a subclass of natural risk statistics including the Basel II and III risk measures (§7).

2. Review of existing risk measures.

2.1. Coherent and convex risk measures.

Artzner et al. [5] propose the *coherent risk measures* that satisfy the following three axioms:

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

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

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

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. [5, p. 209]). We will discuss the controversies related to this axiom in §6. Artzner et al. [5] and Delbaen [11] also present an equivalent approach for defining coherent risk measures via acceptance sets. Föllmer and Schied [14] and Frittelli and Gianin [15] propose the convex risk measures that relax Axioms A1 and A3 to a single convexity axiom: $\rho(\lambda X+(1-\lambda)Y)\leq\lambda\rho(X)+(1-\lambda)\rho(Y)$, $\forall X,Y\in\mathscr{X}$, $\forall\lambda\in[0,1]$.

A risk measure ρ is coherent if and only if there exists a family $\mathscr{Q}$ of probability measures such that $\rho(X)=\sup_{Q\in\mathscr{Q}}\{E^Q[X]\}$, $\forall X\in\mathscr{X}$, where $E^Q[X]$ is the expectation of X under the probability measure Q (see Huber [25], Artzner et al. [5], Delbane [11]). Each $Q\in\mathscr{Q}$ 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., Jaschke and Küchler [30], Staum [45]).

Artzner et al. [5] suggest using a specific risk measure called *tail conditional expectation* (TCE). TCE at level α of X is defined as

$$\mathrm{TCE}_\alpha(X) := \mathrm{E}\big[X \mid X \geq \mathrm{VaR}_\alpha(X)\big]. \tag{1}$$

However, TCE does not generally satisfy subadditivity (see, e.g., Acerbi and Tasche [2, Example 5.4]); hence, the *expected shortfall* (ES) is introduced in Acerbi et al. [1], Tasche [47], and Acerbi and Tasche [2] as a modification of TCE and is shown to be a coherent risk measure. Conditional value-at-risk (CVaR) is introduced in Rockfellar and Uryasev [39] which is equivalent to ES. The ES (or, equivalently, CVaR) at level α of X with the distribution function $F(\cdot)$ is defined to be (Rockfellar and Uryasev [39])

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

where the α-tail distribution of X is defined by the distribution function:

$$F_{\alpha,X}(x) := \begin{cases} 0, & \text{for } x < \mathrm{VaR}_\alpha(X) \\ \dfrac{F(x)-\alpha}{1-\alpha} & \text{for } x \geq \mathrm{VaR}_\alpha(X). \end{cases} \tag{3}$$

If $F(\cdot)$ is continuous, then the α-tail distribution is the same as the distribution of X conditional on that $X \geq \mathrm{VaR}_\alpha(X)$, and $\mathrm{ES}_\alpha(X) = \mathrm{TCE}_\alpha(X)$.

A risk measure is called a *law-invariant* coherent risk measure (Kusuoka [34]) if it satisfies Axioms 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 because 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 §3).

2.2. Insurance risk measures. Wang et al. [49] 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 ω_1 and ω_2 in Ω.)*

Axiom B4. *Continuity*:

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

where $x^+ := \max(x, 0)$, $\forall x \in \mathbb{R}$.

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

Comonotonic random variables are studied by Yaari [50], Schmeidler [41], Denneberg [12], 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 ω changes. For example, the payoffs of a call option and its underlying asset are comonotonic.

Wang et al. [49] show that ρ is an insurance risk measure if and only if ρ has a Choquet integral representation with respect to a distorted probability:

$$\rho(X) = \int X\, d(g\circ P) = \int_{-\infty}^{0} \big(g(P(X>t)) - 1\big)\, dt + \int_0^\infty g\big(P(X>t)\big)\, dt, \tag{4}$$

where $g(\cdot)$ is called the distortion function, which is nondecreasing and satisfies $g(0) = 0$ and $g(1) = 1$. The function $g\circ P$ 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 [12]). 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 §4 for their definition), is not an insurance risk measure, 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. [49] 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 §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 a special case. However, VaR with scenario analysis is 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 (§3.1). (iii) We provide legal and economic reasons for postulating the comonotonic subadditivity axiom (§§5 and 6). (iv) We provide two complete characterizations of the new risk measures (§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 (§§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 $\tilde{x} = (\tilde{x}^1, \tilde{x}^2, \ldots, \tilde{x}^m) \in \mathbb{R}^n$, where $\tilde{x}^i = (x^i_1, \ldots, x^i_{n_i}) \in \mathbb{R}^{n_i}$ is the data subset that corresponds to the ith scenario and n_i is the sample size of $\tilde{x}^i$; $n_1 + n_2 + \cdots + n_m = n$. For each $i = 1, \ldots, m$, $\tilde{x}^i$ 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 §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 that 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. $\tilde{x} = (\tilde{x}^1, \tilde{x}^2, \ldots, \tilde{x}^m) \in \mathbb{R}^n$ and $\tilde{y} = (\tilde{y}^1, \tilde{y}^2, \ldots, \tilde{y}^m) \in \mathbb{R}^n$ are *scenario-wise comonotonic* if for $\forall i$, $\forall 1 \le j, k \le n_i$, it holds that $(x^i_j - x^i_k)(y^i_j - y^i_k) \ge 0$. 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, $i = 1, \ldots, m$, 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*: $\hat{\rho}(a\tilde{x} + b\mathbf{1}) = a\hat{\rho}(\tilde{x}) + sb$, $\forall \tilde{x} \in \mathbb{R}^n$, $\forall a \ge 0$, $\forall b \in \mathbb{R}$, *where* $s > 0$ *is a scaling constant, and* $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$.

Axiom C2. *Monotonicity*: $\hat{\rho}(\tilde{x}) \le \hat{\rho}(\tilde{y})$, *if* $\tilde{x} \le \tilde{y}$, *where* $\tilde{x} \le \tilde{y}$ *means* $x^i_j \le y^i_j$, $j = 1, \ldots, n_i$; $i = 1, \ldots, m$.

These two axioms (with $s=1$ in Axiom C1) are the counterparts of Axioms 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 Axioms 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 $\hat{\rho}(\tilde{x}-\varepsilon\mathbf{1})\le\hat{\rho}(\tilde{y})\le\hat{\rho}(\tilde{x}+\varepsilon\mathbf{1})$. Applying Axiom C1, the inequality further becomes $\hat{\rho}(\tilde{x})-s\varepsilon\le\hat{\rho}(\tilde{y})\le\hat{\rho}(\tilde{x})+s\varepsilon$, which establishes the continuity of $\hat{\rho}$.

Axiom C3. *Scenario-wise comonotonic subadditivity*: $\hat{\rho}(\tilde{x}+\tilde{y})\le\hat{\rho}(\tilde{x})+\hat{\rho}(\tilde{y})$, *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 B1, 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*:

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

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

This axiom is the counterpart of the law invariance Axiom A4. It means that if two data sets $\tilde{x}$ 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 $\hat{\rho}:\mathbb{R}^n\to\mathbb{R}$ is called a *natural risk statistic* if it satisfies Axioms C1–C4. The following theorem completely characterizes natural risk statistics.

Theorem 3.1. (i) *For a given constant* $s>0$ *and a given set of weights* $\mathscr{W}=\{\tilde{w}\}\subset\mathbb{R}^n$ *with each* $\tilde{w}=(w^1_1,\ldots,w^1_{n_1},\ldots,w^m_1,\ldots,w^m_{n_m})\in\mathscr{W}$ *satisfying the following conditions*

$$\sum_{j=1}^{n_1}w^1_j+\sum_{j=1}^{n_2}w^2_j+\cdots+\sum_{j=1}^{n_m}w^m_j=1, \tag{5}$$

$$w^i_j\ge 0,\quad j=1,\ldots,n_i;\, i=1,\ldots,m, \tag{6}$$

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

$$\hat{\rho}(\tilde{x}):=s\cdot\sup_{\tilde{w}\in\mathscr{W}}\left\{\sum_{j=1}^{n_1}w^1_jx^1_{(j)}+\sum_{j=1}^{n_2}w^2_jx^2_{(j)}+\cdots+\sum_{j=1}^{n_m}w^m_jx^m_{(j)}\right\},\quad \forall\tilde{x}=(\tilde{x}^1,\ldots,\tilde{x}^m)\in\mathbb{R}^n, \tag{7}$$

where $(x^i_{(1)},\ldots,x^i_{(n_i)})$ *is the order statistics of* $\tilde{x}^i=(x^i_1,\ldots,x^i_{n_i})$ *with* $x^i_{(n_i)}$ *being the largest*, $i=1,\ldots,m$. *Then the* $\hat{\rho}$ *defined in* (7) *is a natural risk statistic.*

(ii) *If* $\hat{\rho}$ *is a natural risk statistic, then there exists a set of weights* $\mathscr{W}=\{\tilde{w}\}\subset\mathbb{R}^n$ *such that each* $\tilde{w}=(w^1_1,\ldots,w^1_{n_1},\ldots,w^m_1,\ldots,w^m_{n_m})\in\mathscr{W}$ *satisfies condition* (5) *and* (6), *and*

$$\hat{\rho}(\tilde{x})=s\cdot\sup_{\tilde{w}\in\mathscr{W}}\left\{\sum_{j=1}^{n_1}w^1_jx^1_{(j)}+\sum_{j=1}^{n_2}w^2_jx^2_{(j)}+\cdots+\sum_{j=1}^{n_m}w^m_jx^m_{(j)}\right\},\quad \forall\tilde{x}=(\tilde{x}^1,\ldots,\tilde{x}^m)\in\mathbb{R}^n. \tag{8}$$

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 $\mathscr{B}:=\{\tilde{y}=(\tilde{y}^1,\ldots,\tilde{y}^m)\in\mathbb{R}^n\mid y^1_1\le y^1_2\le\cdots\le y^1_{n_1};\ldots;y^m_1\le y^m_2\le\cdots\le y^m_{n_m}\}$. However, unlike the case of coherent risk measures, the existence of a set of weights $\mathscr{W}$ that satisfies (8) does not follow easily from the proof developed by Huber [25]. The main difference here is that the set $\mathscr{B}$ is not an open set in $\mathbb{R}^n$. The boundary points do not have properties as nice 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 et al. [4] provide alternative shorter proofs for Theorems 3.1 and 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 $\hat{\rho}:\mathbb{R}^n \to \mathbb{R}$ is called a *coherent risk statistic* if it satisfies Axioms C1 and C2 and the following Axiom E1.

Axiom E1. *Subadditivity*: $\hat{\rho}(\tilde{x}+\tilde{y}) \le \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y}),\ \forall \tilde{x}, \tilde{y} \in \mathbb{R}^n$.

Theorem 3.2. *A risk statistic $\hat{\rho}$ is a coherent risk statistic if and only if there exists a set of weights $\mathscr{W} = \{\tilde{w}\} \subset \mathbb{R}^n$ such that each $\tilde{w} \in \mathscr{W}$ satisfies (5) and (6), and*

$$\hat{\rho}(\tilde{x}) = s \cdot \sup_{\tilde{w}\in\mathscr{W}} \left\{ \sum_{j=1}^{n_1} w_j^1 x_j^1 + \sum_{j=1}^{n_2} w_j^2 x_j^2 + \cdots + \sum_{j=1}^{n_m} w_j^m x_j^m \right\}, \quad \forall \tilde{x} = (\tilde{x}^1, \ldots, \tilde{x}^m) \in \mathbb{R}^n. \tag{9}$$

Proof. The proof for the "if" part is trivial. To prove the "only if" part, suppose $\hat{\rho}$ is a coherent risk statistic. Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ be a set with n elements and $\mathscr{Z}$ be the set of all real-valued functions defined on Θ. Define the functional $E^*(Z) := (1/s)\hat{\rho}(Z(\theta_1), Z(\theta_2), \ldots, Z(\theta_n)),\ \forall Z \in \mathscr{Z}$. By Axioms C1, C2, and E1, $E^*(\cdot)$ satisfies the conditions in Huber and Ronchetti [26, Proposition 10.1, p. 252], so the result follows by applying that proposition.

Natural risk statistics satisfy empirical law invariance, and 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 $\hat{\rho}$: $\mathbb{R}^n \to \mathbb{R}$ is called an *empirical-law-invariant* coherent risk statistic if it satisfies Axioms C1, C2, C4, and E1. The following theorem completely characterizes empirical-law-invariant coherent risk statistics.

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

$$\sum_{j=1}^{n_1} w_j^1 + \sum_{j=1}^{n_2} w_j^2 + \cdots + \sum_{j=1}^{n_m} w_j^m = 1, \tag{10}$$

$$w_j^i \ge 0, \quad j = 1, \ldots, n_i;\ i = 1, \ldots, m, \tag{11}$$

$$w_1^i \le w_2^i \le \cdots \le w_{n_i}^i, \quad i = 1, \ldots, m, \tag{12}$$

define a risk statistic

$$\hat{\rho}(\tilde{x}) := s \cdot \sup_{\tilde{w}\in\mathscr{W}} \left\{ \sum_{j=1}^{n_1} w_j^1 x_{(j)}^1 + \sum_{j=1}^{n_2} w_j^2 x_{(j)}^2 + \cdots + \sum_{j=1}^{n_m} w_j^m x_{(j)}^m \right\}, \quad \forall \tilde{x} = (\tilde{x}^1, \ldots, \tilde{x}^m) \in \mathbb{R}^n, \tag{13}$$

where $(x_{(1)}^i, \ldots, x_{(n_i)}^i)$ is the order statistics of $\tilde{x}^i = (x_1^i, \ldots, x_{n_i}^i)$ with $x_{(n_i)}^i$ being the largest, $i = 1, \ldots, m$. Then the $\hat{\rho}$ defined in (13) 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 $\mathscr{W} = \{\tilde{w}\} \subset \mathbb{R}^n$ such that each $\tilde{w} \in \mathscr{W}$ satisfies (10), (11), and (12), and*

$$\hat{\rho}(\tilde{x}) = s \cdot \sup_{\tilde{w}\in\mathscr{W}} \left\{ \sum_{j=1}^{n_1} w_j^1 x_{(j)}^1 + \sum_{j=1}^{n_2} w_j^2 x_{(j)}^2 + \cdots + \sum_{j=1}^{n_m} w_j^m x_{(j)}^m \right\}, \quad \forall \tilde{x} = (\tilde{x}^1, \ldots, \tilde{x}^m) \in \mathbb{R}^n. \tag{14}$$

Proof. See Appendix C.

Theorems 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 §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 §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 $\hat{\rho}$: $\mathbb{R}^n \to \mathbb{R}$ is called an *insurance risk statistic* if it satisfies the following Axioms 1–4.

Axiom 1. *Empirical law invariance: the same as Axiom C4.*

Axiom 2. *Monotonicity: $\hat{\rho}(\tilde{x}) \le \hat{\rho}(\tilde{y})$ if $\tilde{x} \le \tilde{y}$.*

Axiom 3. *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 4. *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 $\tilde{w} = (w_1^1, \ldots, w_{n_1}^1, \ldots, w_1^m, \ldots, w_{n_m}^m) \in \mathbb{R}^n$ with $w_j^i \ge 0$ for $j = 1, \ldots, n_i$; $i = 1, \ldots, m$ and $\sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i = 1$, such that*

$$\hat{\rho}(\tilde{x}) = s\left(\sum_{j=1}^{n_1} w_j^1 x_{(j)}^1 + \sum_{j=1}^{n_2} w_j^2 x_{(j)}^2 + \cdots + \sum_{j=1}^{n_m} w_j^m x_{(j)}^m\right), \quad \forall \tilde{x} = (\tilde{x}^1, \tilde{x}^2, \ldots, \tilde{x}^m) \in \mathbb{R}^n, \tag{15}$$

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

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, is not a special case of insurance risk statistics but a special case 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 instance, for a constant $p > 1$, we define the risk measure $\rho_s(X) := \int_{-\infty}^{\infty} |u|^p \, dF_{\alpha,X}(u)$, where $F_{\alpha,X}(\cdot)$ is defined in (3). For X with a continuous distribution, $\rho_s(X)$ is equal to $\mathrm{E}[|X|^p \mid X > \mathrm{VaR}_\alpha(X)]$, which is called the shortfall risk measure in Tasche [46]. Then the risk statistic corresponding to the risk measure ρ_s is not a natural risk statistic because it does not satisfy comonotonic subadditivity. Indeed, in the one-scenario case, for a set of observations $\tilde{x} = (x_1, \ldots, x_n)$ of X, the risk statistic corresponding to ρ_s is defined by $\hat{\rho}_s(\tilde{x}) := \int_{-\infty}^{\infty} |u|^p \, d\hat{F}_{\alpha,X}(u)$, where $\hat{F}_{\alpha,X}(u) := ((F_n(u) - \alpha)/(1-\alpha)) \cdot 1_{\{u \ge x_{(\lceil n\alpha \rceil)}\}}$, $\lceil \cdot \rceil$ is the ceiling function and F_n is the empirical distribution function of X. Then it can be shown that

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

Suppose that $\tilde{x}$ and $\tilde{y} = (y_1, \ldots, y_n)$ are comonotonic, and $x_{(j)} > 0$ and $y_{(j)} > 0$ for all $j \ge k$, then

$$\begin{aligned} \hat{\rho}_s(\tilde{x}+\tilde{y}) &= \frac{k - \alpha n}{(1-\alpha)n} (x_{(k)} + y_{(k)})^p + \frac{1}{(1-\alpha)n} \sum_{j=k+1}^{n} (x_{(j)} + y_{(j)})^p \\ &> \frac{k - \alpha n}{(1-\alpha)n} (x_{(k)}^p + y_{(k)}^p) + \frac{1}{(1-\alpha)n} \sum_{j=k+1}^{n} (x_{(j)}^p + y_{(j)}^p) = \hat{\rho}_s(\tilde{x}) + \hat{\rho}_s(\tilde{y}). \end{aligned}$$

4. Axiomatization of the Basel II and Basel III risk measures. The Basel II Accord [6] 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

$$c_t = \max\left\{\mathrm{VaR}_{t-1}, k \cdot \frac{1}{60} \sum_{i=1}^{60} \mathrm{VaR}_{t-i}\right\}, \tag{16}$$

where k is a constant that is no less than 3; VaR_{t-i} is the 10-day VaR at 99% confidence level calculated on day $t-i$, $i = 1, \ldots, 60$. VaR_{t-i} is usually estimated from a data set $\tilde{x}^i = (x_1^i, x_2^i, \ldots, x_{n_i}^i) \in \mathbb{R}^{n_i}$, which is generated by historical simulation or Monte Carlo simulation (Jorion [31]).

Adrian and Brunnermeier [3] point out that risk measures based on contemporaneous observations, such as the Basel II risk measure (16), are *procyclical*; i.e., risk measurement obtained by such risk measures tends to be low in booms and high in crises, which impedes effective regulation. Gordy and Howells [20] 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 cyclicality 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 that 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 [8], which is defined by

$$c_t = \max\left\{\mathrm{VaR}_{t-1},\, k\cdot\frac{1}{60}\sum_{i=1}^{60}\mathrm{VaR}_{t-i}\right\} + \max\left\{\mathrm{sVaR}_{t-1},\, \ell\cdot\frac{1}{60}\sum_{i=1}^{60}\mathrm{sVaR}_{t-i}\right\}, \tag{17}$$

where VaR_{t-i} is the same as in (16); k and ℓ are constants no less than 3; and 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 (16).

In addition to the capital charge specified in (17), 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 (17) (Basel Committee on Banking Supervision [7, 8]). The IRC capital charge on the tth day is defined as

$$\mathrm{IRC}_t = \max\left\{\mathrm{VaR}^{\mathrm{ir}}_{t-1},\, \frac{1}{60}\sum_{i=1}^{60}\mathrm{VaR}^{\mathrm{ir}}_{t-i}\right\}, \tag{18}$$

where $\mathrm{VaR}^{\mathrm{ir}}_{t-i}$ 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 $\mathrm{VaR}^{\mathrm{ir}}_{t-i}$ 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 [7]). Glasserman [18] 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* (16) *and the Basel III risk measure defined in* (17) *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* (18) *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 procyclical risk measures. Gordy and Howells [20] propose to mitigate the procyclicality of c_t, the Basel II capital requirement, by a method called countercyclical indexing. This applies a time-varying multiplier α_t to c_t and generates a smoothed capital requirement $\alpha_t c_t$, where α_t increases during booms and decreases during recessions to dampen the procyclicality of c_t. In the static setting, the multiplier α_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 countercyclical indexing. Although the current paper focuses on static risk measures, it would be of interest to study axioms for dynamic risk measures that 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 [17], Maccheroni et al. [36], and Hansen and Sargent [21]. 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 [26]. 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 because of the uncertainty in human language and because human beings are unable to anticipate all possible future circumstances (Hart [22, p. 128]). Hence, a law is only a guideline for judges and enforcement officers (Hart [22, 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, the 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 [48, 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 because of 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 because of limited availability of data. For example, the heaviness of tail distributions cannot be identified in many cases. Heyde and Kou [23] 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 is 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. Median shortfall: A robust risk measure. We propose a robust risk measure, *median shortfall* (MS), which is a special case of natural risk statistics. MS is defined by replacing the "mean" in the definition of ES by "median." More precisely, MS of X at level α is defined as

$$\mathrm{MS}_\alpha(X) := \text{median of the } \alpha\text{-tail distribution of } X,$$

where the α-tail distribution of X is defined in (3). It can be shown that for any X, MS at level α of X is equal to VaR of X at level $(1+\alpha)/2$, i.e.,

$$\mathrm{MS}_\alpha(X) = \mathrm{VaR}_{(1+\alpha)/2}(X), \quad \forall X,\ \forall \alpha \in (0,1). \tag{19}$$

Equation (19) 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, MS at 99%, which gives the median of the size of loss beyond 99%. It is also interesting to point out that $\mathrm{MS}_\alpha(X+Y) \le \mathrm{MS}_\alpha(X) + \mathrm{MS}_\alpha(Y)$ may hold for those X and Y that cause $\mathrm{VaR}_\alpha(X+Y) > \mathrm{VaR}_\alpha(X) + \mathrm{VaR}_\alpha(Y)$; in other words, subadditivity may not be violated if one replaces VaR_α by MS_α. Here are two such examples: (i) The example on page 216 of Artzner et al. [5] shows that 99% VaR does not satisfy subadditivity for the two positions of writing an option A and writing an option B. However, 99% MS (or, equivalently, 99.5% VaR) does satisfy subadditivity. Indeed, the 99% MS 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 in Artzner et al. [5, p. 217] shows that the 90% VaR does not satisfy subadditivity for X_1 and X_2. However, the 90% MS (or, equivalently, 95% VaR) does satisfy subadditivity. Actually, the 90% MS of X_1 and X_2 are both equal to 1. By simple calculation, $P(X_1+X_2 \le -2) = 0.005 < 0.05$, which implies that the 90% MS of X_1+X_2 is strictly less than 2.

MS can be shown to be more robust than ES by at least three tools in robust statistics: (i) influence functions, (ii) asymptotic breakdown points, and (iii) finite sample breakdown points. See Appendix F. See also Cont et al. [9] for discussion on robustness of risk measures.

ES is also highly model dependent and particularly sensitive to modeling assumptions on the extreme tails of loss distributions because the computation of ES relies on these extreme tails, as is shown by (F1) in Appendix F. Figure 1 illustrates the sensitivity of ES to modeling assumptions. MS is clearly less sensitive to tail behavior than ES because the changes of MS with respect to the changes of loss distributions have narrower ranges than do those of ES.

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 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 (7) that corresponds to the set of weights $\mathscr{W}$. Define the map ϕ: $\mathscr{W} \to \mathbb{R}^m \times \mathbb{R}^n$ such that $\tilde w \mapsto \phi(\tilde w) := (\tilde p, \tilde q)$, where $\tilde p := (p^1, \ldots, p^m)$, $p^i := \sum_{j=1}^{n_i} w^i_j$, $i = 1, \ldots, m$; $\tilde q := (q^1_1, \ldots, q^1_{n_1}, \ldots, q^m_1, \ldots, q^m_{n_m})$, $q^i_j := 1_{\{p^i>0\}} w^i_j / p^i$. Since $p^i \ge 0$ 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

$$\hat\rho(\tilde x) = s \cdot \sup_{(\tilde p, \tilde q) \in \phi(\mathscr{W})} \left\{ \sum_{i=1}^m p^i \hat\rho^{i,\tilde q}(\tilde x^i) \right\}, \quad \text{where } \hat\rho^{i,\tilde q}(\tilde x^i) := \sum_{j=1}^{n_i} q^i_j x^i_{(j)}. \tag{20}$$

Each weight $\tilde w \in \mathscr{W}$ then corresponds to $\phi(\tilde w) = (\tilde p, \tilde q) \in \phi(\mathscr{W})$, 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, $i = 1, \ldots, m$. 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 statistic.

Example 5.1. MS (or, equivalently, VaR at a higher confidence level) is a robust statistic. Another example of robust statistics is the sample version of the following new risk measure which we call *trimmed average* VaR (tav):

$$\rho_{\text{tav}}(X) := \frac{1}{\beta-\alpha} \int_\alpha^\beta F^{-1}(u)\, du, \tag{21}$$

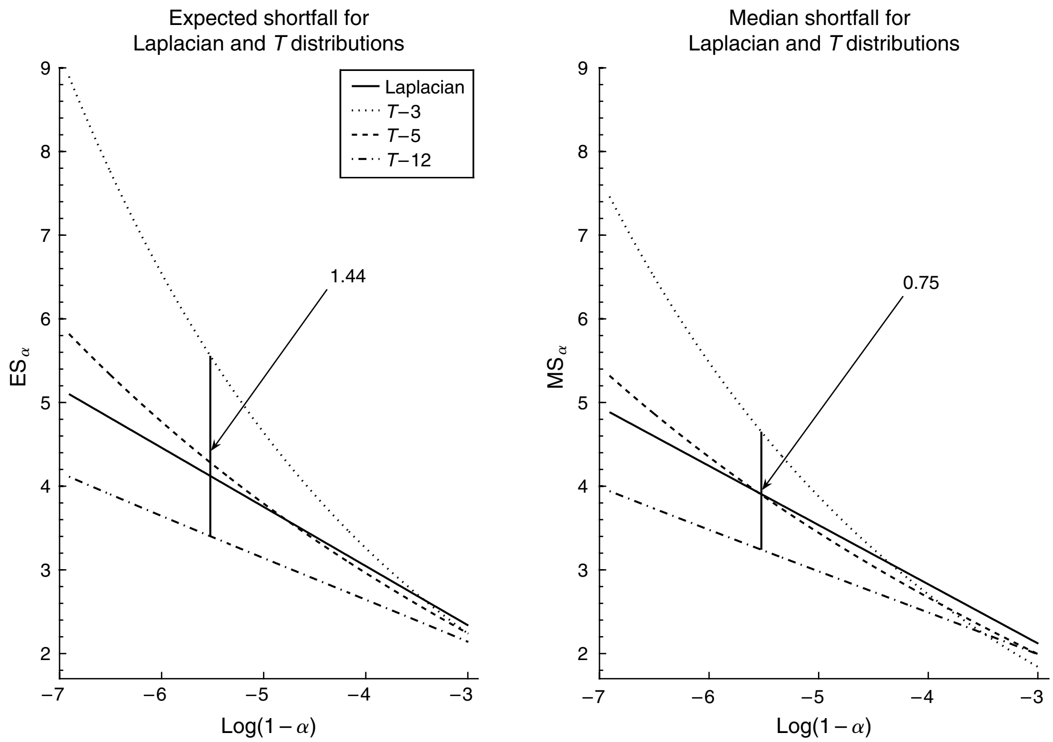

Figure 1. Comparing the robustness of expected shortfall (ES) and median shortfall (MS) with respect to model misspecification. ES_α and MS_α are calculated for Laplace and T distributions with degree of freedom 3, 5, and 12, which are normalized to have mean 0 and variance 1. The horizontal axis is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

where $0 < \alpha < \beta < 1$, e.g., $\alpha = 99\%$, $\beta = 99.9\%$. ρ_{tav} is robust because it does not involve quantiles with levels higher than β. It can be shown that the sample version of ρ_{tav} corresponding to the data $\tilde{x}^i$ is given by

$$\hat{\rho}_{\mathrm{tav}}(\tilde{x}^i) = \frac{1}{\beta-\alpha}\left(\frac{k_{1,i}-n_i\alpha}{n_i}x^i_{(k_{1,i})} + \sum_{j=k_{1,i}+1}^{k_{2,i}-1}\frac{1}{n_i}x^i_{(j)} + \frac{1+n_i\beta-k_{2,i}}{n_i}x^i_{(k_{2,i})}\right), \qquad k_{1,i} = \lceil n_i\alpha\rceil, \quad k_{2,i} = \lceil n_i\beta\rceil.$$

Example 5.2. The Basel II risk measure (16) 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 $1/k$ to the scenario of day $t-1$ and $1-1/k$ to an imaginary scenario under which losses are identically 0; the other prior assigns probability $1/60$ to each of the scenarios corresponding to day $t-i$, $i = 1, \ldots, 60$.

Example 5.3. The Basel III risk measure (17) is more robust than is the Basel II risk measure (16) 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 (16), the Basel III IRC risk measure (18) is robust in the sense that each subsidiary risk statistic $\mathrm{VaR}^{\mathrm{ir}}_{t-i}$ is robust, and the risk measure incorporates two priors of probability distributions on the set of scenarios.

5.6. Neither law-invariant coherent risk measures nor insurance risk measures are robust. No law-invariant coherent risk measure is robust with respect to small changes in the data. Indeed, by Theorem 3.3, an empirical-law-invariant coherent risk statistic $\hat{\rho}$ can be represented by (14), where for each weight $\tilde{w}$, w^i_j 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 $\max\{x^i_{(n_i)}: i = 1, \ldots, m\}$, which is not robust at all. In general, the finite sample breakdown point (see Huber and Ronchetti [26, 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, ES is sensitive to modeling assumptions of heaviness of tail distributions and to outliers in the data, as is shown in §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 because it generally increases the safety of the financial system. Of course, risk measures that 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 ES is more conservative than is MS because the median can be bigger than the mean for some distributions.

A natural risk statistic can be constructed by (7) in the following ways so that it is both conservative and robust: (i) more data subsets that correspond to stressed scenarios can be included in (7), and (ii) a larger constant s in (7) can be used. For example, adding 60 stressed scenarios makes (17) much more conservative than is (16), and a larger k or ℓ in (17) 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 [40] 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 [37], Hong and Herk [24], and Kijima [33] 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 [27, 28] and Ibragimov and Walden [29] 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, and (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 nondiversification, 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. [5, pp. 217–218] show that VaR prohibits diversification by a simple example 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 [27, 28] and Ibragimov and Walden [29] 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, Daníelsson et al. [10] 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. Daníelsson et al. [10] 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 et al. [16] 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 Daníelsson et al. [10] also show that VaR_α 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. [5, p. 209]). However, Dhaene et al. [13] 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 that 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 [35]).

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 et al. [38]). Consider a portfolio composed of u_i units of asset i, $i = 1, \ldots, d$, and denote $u = (u_1, u_2, \ldots, u_d)$. Suppose that there are m scenarios. Let $\tilde{x}(i) = (\tilde{x}(i)^1, \tilde{x}(i)^2, \ldots, \tilde{x}(i)^m)$ be the observed loss of the ith asset, where $\tilde{x}(i)^s = (x(i)^s_1, x(i)^s_2, \ldots, x(i)^s_{n_s}) \in \mathbb{R}^{n_s}$ are the observations under the sth scenario, $s = 1, \ldots, m$. Then the observations of the portfolio loss are given by $\tilde{l}(u) = \sum_{i=1}^d u_i \tilde{x}(i) = (\tilde{l}(u)^1, \tilde{l}(u)^2, \ldots, \tilde{l}(u)^m)$, where $\tilde{l}(u)^s = (l(u)^s_1, l(u)^s_2, \ldots, l(u)^s_{n_s}) \in \mathbb{R}^{n_s}$ and $l(u)^s_j := \sum_{i=1}^d u_i x(i)^s_j$. 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

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

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

Proposition 7.1. *Suppose $(\tilde{x}(1), \tilde{x}(2), \ldots, \tilde{x}(d))$ is a sample of the random vector $(X(1), X(2), \ldots, X(d))$, where $X(i) = (X(i)^1, X(i)^2, \ldots, X(i)^m)$ and $X(i)^s = (X(i)^s_1, X(i)^s_2, \ldots, X(i)^s_{n_s}) \in \mathbb{R}^{n_s}$. Suppose that the joint distribution of $(X(1), X(2), \ldots, X(d))$ has a probability density on $\mathbb{R}^{dn}$. Then for any given $u \neq 0$, it holds with probability 1 that*

$$C_{\hat{\rho}}(u) = \sum_{i=1}^d u_i \frac{\partial C_{\hat{\rho}}(u)}{\partial u_i}, \tag{22}$$

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

Proof. For any given $u \neq 0$, let $\mathbb{X}_u$ be the set of samples $(\tilde{x}(1), \tilde{x}(2), \ldots, \tilde{x}(d)) \in \mathbb{R}^{dn}$ that satisfies the following conditions: (i) $l(u)^1_{(\lceil \alpha n_1 \rceil)} \neq (k/60) \sum_{s=1}^{60} l(u)^s_{(\lceil \alpha n_s \rceil)}$; (ii) $l(u)^{61}_{(\lceil \alpha n_{61} \rceil)} \neq (\ell/60) \sum_{s=61}^{120} l(u)^s_{(\lceil \alpha n_s \rceil)}$; (iii) $l(u)^s_i \neq l(u)^s_j$ for any s and $i \neq j$. Then it follows from the condition of the proposition that $P((X(1), X(2), \ldots, X(d)) \in \mathbb{X}_u) = 1$. Fix any $(\tilde{x}(1), \tilde{x}(2), \ldots, \tilde{x}(d)) \in \mathbb{X}_u$. 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 $V := \{v \in \mathbb{R}^d \mid \|v - u\| < \delta\}$. Hence, $C_{\hat{\rho}}(\cdot)$ is differentiable at u, and (22) holds.

For any given $u \neq 0$, let $\mathbb{X}_u$ be defined in the above proof and suppose $\tilde{x} \in \mathbb{X}_u$. To compute $u_i(\partial C_{\hat{\rho}}(u)/\partial u_i)$, one only needs to compute $(\partial l(u)^s_{(\lceil \alpha n_s \rceil)})/\partial u_i$. Let $(p_1, \ldots, p_{n_s})$ be the permutation of $(1, 2, \ldots, n_s)$ such that $l(u)^s_{p_1} < l(u)^s_{p_2} < \cdots < l(u)^s_{p_{n_s}}$. Then there exists $\delta > 0$ such that $l(v)^s_{p_1} < l(v)^s_{p_2} < \cdots < l(v)^s_{p_{n_s}}$ for $\forall v \in V$, where $V := \{v \in \mathbb{R}^d \mid \|v - u\| < \delta\}$. Hence, for $\forall v \in V$,

$$l(v)^s_{(\lceil \alpha n_s \rceil)} = l(v)^s_{p_{\lceil \alpha n_s \rceil}} = \sum_{i=1}^d v_i x(i)^s_{p_{\lceil \alpha n_s \rceil}}, \qquad \text{and} \qquad \frac{\partial l(u)^s_{(\lceil \alpha n_s \rceil)}}{\partial u_i} = x(i)^s_{p_{\lceil \alpha n_s \rceil}}.$$

In general, let Υ_1 be the set of natural risk statistic $\hat{\rho}$ that can be represented by (8) using only a finite set $\mathscr{W}$. Let Υ_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 $\hat{\rho}_k \in \Upsilon_1$, $k = 1, \ldots, K$. Both the Basel II risk measure and Basel III risk measure belong to the set Υ_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 piecewise 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 countercyclical indexing risk measure suggested by Gordy and Howells [20] 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 data-based law-invariant coherent risk measures are generally not robust with respect to small changes in the data and data-based 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 median shortfall (with scenario analysis), which is more robust than expected shortfall suggested by the theory of coherent risk measures. The Euler capital allocation rule 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

$$\mathscr{B} := \{\tilde{y} = (\tilde{y}^1, \ldots, \tilde{y}^m) \in \mathbb{R}^n \mid y_1^1 \leq y_2^1 \leq \cdots \leq y_{n_1}^1; \ldots; y_1^m \leq y_2^m \leq \cdots \leq y_{n_m}^m\}. \tag{A1}$$

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 $\mathscr{B}$ be defined in (A1) and $\mathscr{B}^o$ be the interior of $\mathscr{B}$. For any fixed $\tilde{z} \in \mathscr{B}^o$ and any $\hat{\rho}$ satisfying Axioms C1–C4 and $\hat{\rho}(\tilde{z}) = 1$, there exists a weight $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^m) \in \mathbb{R}^n$ such that the linear functional $\lambda(\tilde{x}) := \sum_{j=1}^{n_1} w_j^1 x_j^1 + \sum_{j=1}^{n_2} w_j^2 x_j^2 + \cdots + \sum_{j=1}^{n_m} w_j^m x_j^m$ satisfies*

$$\lambda(\tilde{z}) = 1, \tag{A2}$$

$$\lambda(\tilde{x}) < 1, \quad \text{for any } \tilde{x} \text{ such that } \tilde{x} \in \mathscr{B} \text{ and } \hat{\rho}(\tilde{x}) < 1. \tag{A3}$$

Proof. Let $U = \{\tilde{x} = (\tilde{x}^1, \ldots, \tilde{x}^m) \mid \hat{\rho}(\tilde{x}) < 1\} \cap \mathscr{B}$. For any $\tilde{x} = (\tilde{x}^1, \ldots, \tilde{x}^m) \in \mathscr{B}$ and $\tilde{y} = (\tilde{y}^1, \ldots, \tilde{y}^m) \in \mathscr{B}$, $\tilde{x}$ and $\tilde{y}$ are scenario-wise comonotonic. Then Axioms 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 $\tilde{z} - \varepsilon\mathbf{1} \in U$. Because $\tilde{z} - \varepsilon\mathbf{1}$ 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 $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^m) = (w_1^1, \ldots, w_{n_1}^1, \ldots, w_1^m, \ldots, w_{n_m}^m) \in \mathbb{R}^n$ such that $\lambda(\tilde{x}) := \sum_{j=1}^{n_1} w_j^1 x_j^1 + \sum_{j=1}^{n_2} w_j^2 x_j^2 + \cdots + \sum_{j=1}^{n_m} w_j^m x_j^m$ satisfies $\lambda(\tilde{x}) \leq \lambda(\tilde{z})$ for any $\tilde{x} \in \bar{U}$. In particular, we have

$$\lambda(\tilde{x}) \leq \lambda(\tilde{z}), \forall \tilde{x} \in U. \tag{A4}$$

We shall show that the strict inequality holds in (A4). 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

$$\lambda(\alpha\tilde{z} + (1-\alpha)\tilde{r}) = \alpha\lambda(\tilde{z}) + (1-\alpha)\lambda(\tilde{r}) = \lambda(\tilde{z}). \tag{A5}$$

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

$$\hat{\rho}(\alpha\tilde{z} + (1-\alpha)\tilde{r}) \leq \alpha\hat{\rho}(\tilde{z}) + (1-\alpha)\hat{\rho}(\tilde{r}) < \alpha + (1-\alpha) = 1, \quad \forall \alpha \in (0, 1). \tag{A6}$$

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

$$\alpha_0 \tilde{z} + (1-\alpha_0)\tilde{r} + \varepsilon \tilde{w} \in \mathscr{B}. \tag{A7}$$

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

$$\hat{\rho}\big(\alpha_0\tilde{z} + (1-\alpha_0)\tilde{r} + \varepsilon\tilde{w}\big) \le \hat{\rho}\big(\alpha_0\tilde{z} + (1-\alpha_0)\tilde{r} + \varepsilon w_{\max}\mathbf{1}\big) = \hat{\rho}\big(\alpha_0\tilde{z} + (1-\alpha_0)\tilde{r}\big) + \varepsilon w_{\max}. \tag{A8}$$

Since $\hat{\rho}(\alpha_0\tilde{z} + (1-\alpha_0)\tilde{r}) < 1$ via (A6), we have by (A8) and (A7) 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, (A4) implies $\lambda(\alpha_0\tilde{z} + (1-\alpha_0)\tilde{r} + \varepsilon\tilde{w}) \le \lambda(\tilde{z})$. However, we have, by (A5), 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

$$\lambda(\tilde{x}) < \lambda(\tilde{z}), \forall \tilde{x} \in U. \tag{A9}$$

Since $\hat{\rho}(0) = 0$, we have $0 \in U$. Letting $\tilde{x} = 0$ in (A9) yields $\lambda(\tilde{z}) > 0$, so we can rescale $\tilde{w}$ such that $\lambda(\tilde{z}) = 1 = \hat{\rho}(\tilde{z})$. Thus, (A9) becomes $\lambda(\tilde{x}) < 1$ for any $\tilde{x}$ such that $\tilde{x} \in \mathscr{B}$ and $\hat{\rho}(\tilde{x}) < 1$, from which (A3) holds.

Lemma A.2. *Let $\mathscr{B}$ be defined in (A1) and $\mathscr{B}^o$ be the interior of $\mathscr{B}$. For any fixed $\tilde{z} \in \mathscr{B}^o$ and any $\hat{\rho}$ satisfying Axioms C1–C4, there exists a weight $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^m) \in \mathbb{R}^n$ such that $\tilde{w}$ satisfies (5) and (6), and*

$$\hat{\rho}(\tilde{x}) \ge \sum_{i=1}^{m}\sum_{j=1}^{n_i} w_j^i x_j^i \quad \text{for any } \tilde{x} \in \mathscr{B}, \qquad \text{and} \quad \hat{\rho}(\tilde{z}) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} w_j^i z_j^i. \tag{A10}$$

Proof. We will show this by considering three cases.

Case 1. $\hat{\rho}(\tilde{z}) = 1$. From Lemma A.1, there exists a weight $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^m) \in \mathbb{R}^n$ such that the linear functional $\lambda(\tilde{x}) := \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i$ satisfies (A2) and (A3).

Firstly, we prove that $\tilde{w}$ satisfies (5), 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, (A3) implies $\lambda(c\mathbf{1}) < 1$, and, by continuity of $\lambda(\cdot)$, we obtain that $\lambda(\mathbf{1}) \le 1$. 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 (A3) and (A2) 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 $\lambda(\mathbf{1}) \ge 1$, and $\tilde{w}$ satisfies (5).

Secondly, we prove that $\tilde{w}$ satisfies (6). For any fixed i and $1 \le j \le n_i$, let $k = n_1 + n_2 + \cdots + n_{i-1} + j$ and $\tilde{e} = (0, \ldots, 0, 1, 0, \ldots, 0)$ be the kth standard basis of $\mathbb{R}^n$. Then $w_j^i = \lambda(\tilde{e})$. Since $\tilde{z} \in \mathscr{B}^o$, there exists $\delta > 0$ such that$\tilde{z} - \delta\tilde{e} \in \mathscr{B}$. For any $\varepsilon > 0$, Axioms C1 and C2 imply $\hat{\rho}(\tilde{z} - \delta\tilde{e} - \varepsilon\mathbf{1}) = \hat{\rho}(\tilde{z} - \delta\tilde{e}) - \varepsilon \le \hat{\rho}(\tilde{z}) - \varepsilon = 1 - \varepsilon < 1$. Then (A3) and (A2) 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 ε go to 0.

Thirdly, we prove that $\tilde{w}$ satisfies (A10). It follows from Axiom C1 and (A3) that

$$\forall c > 0,\ \lambda(\tilde{x}) < c \quad \text{for any } \tilde{x} \text{ such that } \tilde{x} \in \mathscr{B} \text{ and } \hat{\rho}(\tilde{x}) < c. \tag{A11}$$

For any $c \le 0$, we choose $b > 0$ such that $b + c > 0$. Then by (A11), we have $\lambda(\tilde{x} + b\mathbf{1}) < c + b$ for any $\tilde{x}$ such that $\tilde{x} \in \mathscr{B}$ 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

$$\forall c \le 0,\ \lambda(\tilde{x}) < c \quad \text{for any } \tilde{x} \text{ such that } \tilde{x} \in \mathscr{B} \text{ and } \hat{\rho}(\tilde{x}) < c. \tag{A12}$$

It follows from (A11) and (A12) that $\hat{\rho}(\tilde{x}) \ge \lambda(\tilde{x})$ for any $\tilde{x} \in \mathscr{B}$, which in combination with $\hat{\rho}(\tilde{z}) = 1 = \lambda(\tilde{z})$ completes the proof of (A10).

Case 2. $\hat{\rho}(\tilde{z}) \ne 1$ and $\hat{\rho}(\tilde{z}) > 0$. Since $\hat{\rho}((1/\hat{\rho}(\tilde{z}))\tilde{z}) = 1$ and $(1/\hat{\rho}(\tilde{z}))\tilde{z} \in \mathscr{B}^o$, it follows from the result proved in Case 1 that there exists a weight $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^m) \in \mathbb{R}^n$ such that $\tilde{w}$ satisfies (5), (6), and the linear functional $\lambda(\tilde{x}) := \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i$ satisfies $\hat{\rho}(\tilde{x}) \ge \lambda(\tilde{x})$ for $\forall \tilde{x} \in \mathscr{B}$ and $\hat{\rho}((1/\hat{\rho}(\tilde{z}))\tilde{z}) = \lambda((1/\hat{\rho}(\tilde{z}))\tilde{z})$, or, equivalently, $\hat{\rho}(\tilde{z}) = \lambda(\tilde{z})$. Thus, $\tilde{w}$ also satisfies (A10).

Case 3. $\hat{\rho}(\tilde{z}) \le 0$. Choose $b > 0$ such that $\hat{\rho}(\tilde{z} + b\mathbf{1}) > 0$. Since $\tilde{z} + b\mathbf{1} \in \mathscr{B}^o$, it follows from the results proved in Case 1 and Case 2 that there exists a weight $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^m) \in \mathbb{R}^n$ such that $\tilde{w}$ satisfies (5) and (6), and the linear functional $\lambda(\tilde{x}) := \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i$ satisfies $\hat{\rho}(\tilde{x}) \ge \lambda(\tilde{x})$ for $\forall \tilde{x} \in \mathscr{B}$, 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 (A10).

Proof of Theorem 3.1 Firstly, we prove part (i). Suppose $\hat{\rho}$ is defined by (7); then obviously $\hat{\rho}$ satisfies Axioms C1 and C4. To check Axiom C2, suppose $\tilde{x} \le \tilde{y}$. For each $i = 1, \ldots, m$, let $(p_{i,1}, \ldots, p_{i,n_i})$ be the permutation of $(1, \ldots, n_i)$ such that $(y_{(1)}^i, y_{(2)}^i, \ldots, y_{(n_i)}^i) = (y_{p_{i,1}}^i, y_{p_{i,2}}^i, \ldots, y_{p_{i,n_i}}^i)$. Then for any $1 \le j \le n_i$ and $1 \le i \le m$, $y_{(j)}^i = y_{p_{i,j}}^i = \max\{y_{p_{i,k}}^i; k = 1, \ldots, j\} \ge \max\{x_{p_{i,k}}^i; k = 1, \ldots, j\} \ge x_{(j)}^i$, which implies that $\hat{\rho}$ satisfies Axiom C2 because

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

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

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

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

Secondly, we prove part (ii). Let $\mathscr{B}$ be defined in (A1). By Axiom C4, we only need to show that there exists a set of weights $\mathscr{W} = \{\tilde{w}\} \subset \mathbb{R}^n$ such that each $\tilde{w} \in \mathscr{W}$ satisfies condition (5) and (6), and

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathscr{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i \right\}$$

for $\forall \tilde{x} \in \mathscr{B}$.

By Lemma A.2, for any point $\tilde{y} \in \mathscr{B}^o$, there exists a weight $\widetilde{w(\tilde{y})} = (w(\tilde{y})_1^1, \ldots, w(\tilde{y})_{n_1}^1; \ldots; w(\tilde{y})_1^m, \ldots, w(\tilde{y})_{n_m}^m) \in \mathbb{R}^n$ such that (5) and (6) hold and that

$$\hat{\rho}(\tilde{x}) \ge \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i x_j^i \quad \text{for } \forall \tilde{x} \in \mathscr{B} \qquad \text{and} \qquad \hat{\rho}(\tilde{y}) = \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i y_j^i. \tag{A13}$$

Define $\mathscr{W}$ as the collection of such weights; i.e., $\mathscr{W} := \{\widetilde{w(\tilde{y})} \mid \tilde{y} \in \mathscr{B}^o\}$, then each $\tilde{w} \in \mathscr{W}$ satisfies (5) and (6). From (A13), for any fixed $\tilde{x} \in \mathscr{B}^o$, we have

$$\hat{\rho}(\tilde{x}) \ge \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i x_j^i \quad \text{for } \forall \tilde{y} \in \mathscr{B}^o \qquad \text{and} \qquad \hat{\rho}(\tilde{x}) = \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{x})_j^i x_j^i.$$

Therefore,

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{y} \in \mathscr{B}^o} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w(\tilde{y})_j^i x_j^i \right\} = \sup_{\tilde{w} \in \mathscr{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i \right\}, \quad \forall \tilde{x} \in \mathscr{B}^o. \tag{A14}$$

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

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathscr{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i x_j^i \right\}, \quad \forall \tilde{x} \in \partial\mathscr{B}. \tag{A15}$$

Let $\tilde{b} = (b_1^1, \ldots, b_{n_1}^1, \ldots, b_1^m, \ldots, b_{n_m}^m)$ be any boundary point of $\mathscr{B}$. Then there exists a sequence $\{\widetilde{b(k)}\}_{k=1}^\infty \subset \mathscr{B}^o$ such that $\widetilde{b(k)} \to \tilde{b}$ as $k \to \infty$. By the continuity of $\hat{\rho}$ and (A14), we have

$$\hat{\rho}(\tilde{b}) = \lim_{k \to \infty} \hat{\rho}(\widetilde{b(k)}) = \lim_{k \to \infty} \sup_{\tilde{w} \in \mathscr{W}} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} w_j^i b(k)_j^i \right\}. \tag{A16}$$

If we can interchange sup and limit in (A16)—i.e. if

$$\lim_{k\to\infty}\sup_{\tilde{w}\in\mathcal{W}}\left\{\sum_{i=1}^{m}\sum_{j=1}^{n_i}w_j^i b(k)_j^i\right\}=\sup_{\tilde{w}\in\mathcal{W}}\left\{\lim_{k\to\infty}\sum_{i=1}^{m}\sum_{j=1}^{n_i}w_j^i b(k)_j^i\right\}=\sup_{\tilde{w}\in\mathcal{W}}\left\{\sum_{i=1}^{m}\sum_{j=1}^{n_i}w_j^i b_j^i\right\},\tag{A17}$$

—then (A15) holds and the proof is completed. To show (A17), note by Cauchy-Schwarz inequality

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

because $w_j^i\ge 0$ and $\sum_{i=1}^m\sum_{j=1}^{n_i}w_j^i=1, \forall\tilde{w}\in\mathcal{W}$. Hence, $\sum_{i=1}^m\sum_{j=1}^{n_i}w_j^i b(k)_j^i\to\sum_{i=1}^m\sum_{j=1}^{n_i}w_j^i b_j^i$ uniformly for all $\tilde{w}\in\mathcal{W}$ as $k\to\infty$. Therefore, (A17) follows.

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

$$\hat{\rho}_{\mathcal{A}}(\tilde{x}):=\inf\{h\mid \tilde{x}-h\mathbf{1}\in\mathcal{A}\},\ \ \forall\tilde{x}\in\mathbb{R}^n.\tag{B1}$$

$\hat{\rho}_{\mathcal{A}}(\tilde{x})$ 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

$$\mathcal{A}_{\hat{\rho}}:=\{\tilde{x}\in\mathbb{R}^n\mid\hat{\rho}(\tilde{x})\le 0\}.\tag{B2}$$

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

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

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: if $\tilde{x}\in\mathcal{A}$, then $\lambda\tilde{x}\in\mathcal{A}$ for any $\lambda\ge 0$.*

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

Axiom D6. *$\mathcal{A}$ is empirical-law-invariant: 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 $(x^1_{p_{1,1}},x^1_{p_{1,2}},\dots,x^1_{p_{1,n_1}},\dots,x^m_{p_{m,1}},x^m_{p_{m,2}},\dots,x^m_{p_{m,n_m}})\in\mathcal{A}$.*

The following theorem shows that a natural risk statistic and a statistical acceptance set satisfying Axioms 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 Axioms D1–D6.*

(ii) *If a statistical acceptance set $\mathcal{A}$ satisfies Axioms D1–D6, then the risk statistic $\hat{\rho}_{\mathcal{A}}$ is a natural risk statistic (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 Axioms D1–D6, then $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}}=\bar{\mathcal{D}}$, the closure of $\mathcal{D}$.*

Proof. (i) (1) For $\forall\tilde{x}\le 0$, Axiom C2 implies $\hat{\rho}(\tilde{x})\le\hat{\rho}(0)=0$. Hence, $\tilde{x}\in\mathcal{A}_{\hat{\rho}}$ by definition. Thus, D1 holds. (2) For any $\tilde{x}\in\mathbb{R}^n_{++}$, there exists $\alpha>0$ such that $0\le\tilde{x}-\alpha\mathbf{1}$. Axioms C1 and C2 imply that $\hat{\rho}(0)\le\hat{\rho}(\tilde{x}-\alpha\mathbf{1})=\hat{\rho}(\tilde{x})-s\alpha$. So $\hat{\rho}(\tilde{x})\ge s\alpha>0$ 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 $\hat{\rho}(\tilde{x})\le 0$, $\hat{\rho}(\tilde{y})\le 0$, and $\lambda\tilde{x}$ and $(1-\lambda)\tilde{y}$ are scenario-wise comonotonic for any $\lambda\in[0,1]$. Thus, Axiom C3 implies $\hat{\rho}(\lambda\tilde{x}+(1-\lambda)\tilde{y})\le\hat{\rho}(\lambda\tilde{x})+\hat{\rho}((1-\lambda)\tilde{y})=\lambda\hat{\rho}(\tilde{x})+(1-\lambda)\hat{\rho}(\tilde{y})\le 0$. Hence, $\lambda\tilde{x}+(1-\lambda)\tilde{y}\in\mathcal{A}_{\hat{\rho}}$; i.e., D3 holds. (4) For any $\tilde{x}\in\mathcal{A}_{\hat{\rho}}$ and $a>0$, we have $\hat{\rho}(\tilde{x})\le 0$ and Axiom C1 implies $\hat{\rho}(a\tilde{x})=a\hat{\rho}(\tilde{x})\le 0$. Thus, $a\tilde{x}\in\mathcal{A}_{\hat{\rho}}$; i.e., D4 holds. (5) For any $\tilde{x}\le\tilde{y}$ and $\tilde{y}\in\mathcal{A}_{\hat{\rho}}$, we have $\hat{\rho}(\tilde{y})\le 0$.

By Axiom C2, $\hat{\rho}(\tilde{x}) \le \hat{\rho}(\tilde{y}) \le 0$. Hence, $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$; i.e., D5 holds. (6) If $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$, then $\hat{\rho}(\tilde{x}) \le 0$. For any permutation $(p_{i,1}, p_{i,2}, \ldots, p_{i,n_i})$ of $(1, 2, \ldots, n_i)$, $i = 1, \ldots, m$, Axiom C4 implies $\hat{\rho}((x^1_{p_{1,1}}, x^1_{p_{1,2}}, \ldots, x^1_{p_{1,n_1}}, \ldots, x^m_{p_{m,1}}, x^m_{p_{m,2}}, \ldots, x^m_{p_{m,n_m}})) = \hat{\rho}(\tilde{x}) \le 0$. So $(x^1_{p_{1,1}}, x^1_{p_{1,2}}, \ldots, x^1_{p_{1,n_1}}, \ldots, x^m_{p_{m,1}}, x^m_{p_{m,2}}, \ldots, x^m_{p_{m,n_m}}) \in \mathcal{A}_{\hat{\rho}}$; i.e., D6 holds. (7) Suppose $\{\widetilde{x(k)}\}_{k=1}^{\infty} \subset \mathcal{A}_{\hat{\rho}}$, and $\widetilde{x(k)} \to \tilde{x}$ as $k \to \infty$. Then $\hat{\rho}(\widetilde{x(k)}) \le 0$, $\forall k$. The continuity of $\hat{\rho}$ (see the comment following the definition of Axiom C2) implies $\hat{\rho}(\tilde{x}) = \lim_{k\to\infty} \hat{\rho}(\widetilde{x(k)}) \le 0$. So $\tilde{x} \in \mathcal{A}_{\hat{\rho}}$; i.e., $\mathcal{A}_{\hat{\rho}}$ is closed.

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

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

For $\forall \tilde{x} \in \mathbb{R}^n, \forall a \ge 0$, 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 Axioms 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, Axiom C1 holds (with $s = 1$). (2) Suppose $\tilde{x} \le \tilde{y}$. For any $h \in \mathbb{R}$, if $\tilde{y} - h\mathbf{1} \in \mathcal{A}$, then Axiom D5 and $\tilde{x} - h\mathbf{1} \le \tilde{y} - h\mathbf{1}$ 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 $\hat{\rho}_{\mathcal{A}}(\tilde{y}) \ge \hat{\rho}_{\mathcal{A}}(\tilde{x})$; 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 $\tilde{y} - h\mathbf{1} \in \mathcal{A}$, because $\tilde{x} - g\mathbf{1}$ 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, $\hat{\rho}_{\mathcal{A}}(\tilde{x} + \tilde{y}) \le g + h$. 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 $\hat{\rho}_{\mathcal{A}}(\tilde{x} + \tilde{y}) \le \hat{\rho}_{\mathcal{A}}(\tilde{x}) + \hat{\rho}_{\mathcal{A}}(\tilde{y})$. So C3 holds. (4) Fix any $\tilde{x} \in \mathbb{R}^n$ and any permutation $(p_{i,1}, p_{i,2}, \ldots, p_{i,n_i})$ of $(1, 2, \ldots, n_i)$, $i = 1, \ldots, m$. Then for any $h \in \mathbb{R}$, Axiom D6 implies that $\tilde{x} - h\mathbf{1} \in \mathcal{A}$ if and only if $(x^1_{p_{1,1}}, x^1_{p_{1,2}}, \ldots, x^1_{p_{1,n_1}}, \ldots, x^m_{p_{m,1}}, x^m_{p_{m,2}}, \ldots, x^m_{p_{m,n_m}}) - h\mathbf{1} \in \mathcal{A}$. Hence, $\{h \mid \tilde{x} - h\mathbf{1} \in \mathcal{A}\} = \{h \mid (x^1_{p_{1,1}}, x^1_{p_{1,2}}, \ldots, x^1_{p_{1,n_1}}, \ldots, x^m_{p_{m,1}}, x^m_{p_{m,2}}, \ldots, x^m_{p_{m,n_m}}) - h\mathbf{1} \in \mathcal{A}\}$. Taking infimum, we obtain $\hat{\rho}_{\mathcal{A}}(\tilde{x}) = \hat{\rho}_{\mathcal{A}}((x^1_{p_{1,1}}, x^1_{p_{1,2}}, \ldots, x^1_{p_{1,n_1}}, \ldots, x^m_{p_{m,1}}, x^m_{p_{m,2}}, \ldots, x^m_{p_{m,n_m}}))$; i.e., C4 holds.

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

(iv) For any $\tilde{x} \in \mathcal{D}$, we have $\hat{\rho}_{\mathcal{D}}(\tilde{x}) \le 0$. 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 $\hat{\rho}_{\mathcal{D}}(\tilde{x}) \le 0$; i.e., $\inf\{h \mid \tilde{x} - h\mathbf{1} \in \mathcal{D}\} \le 0$. 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 $k \to \infty$ 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 $\mathcal{A}_{\hat{\rho}_{\mathcal{D}}} = \bar{\mathcal{D}}$.

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 (A1). For any fixed $\tilde{z} \in \mathcal{B}$ and any $\hat{\rho}$ satisfying Axioms C1–C2, C4, and E1, and $\hat{\rho}(\tilde{z}) = 1$, there exists a weight $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^m) \in \mathbb{R}^n$ satisfying (12) such that the linear functional $\lambda(\tilde{x}) := \sum_{j=1}^{n_1} w^1_j x^1_j + \sum_{j=1}^{n_2} w^2_j x^2_j + \cdots + \sum_{j=1}^{n_m} w^m_j x^m_j$ satisfies*

$$\lambda(\tilde{z}) = 1, \tag{C1}$$

$$\lambda(\tilde{x}) < 1 \quad \textit{for any } \tilde{x} \textit{ such that } \ \hat{\rho}(\tilde{x}) < 1. \tag{C2}$$

Proof. Let $U = \{\tilde{x} \mid \hat{\rho}(\tilde{x}) < 1\}$. Axioms C1 and E1 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 $\tilde{z} - \varepsilon\mathbf{1} \in U$. Because $\tilde{z} - \varepsilon\mathbf{1}$ converges to $\tilde{z}$ as $\varepsilon \downarrow 0$ and $\hat{\rho}(\tilde{z}) = 1$, $\tilde{z}$ is a boundary point of U. Therefore, there exists a supporting hyperplane for $\bar{U}$ at $\tilde{z}$; i.e., there exists a nonzero vector $\tilde{u} = (u^1_1, \ldots, u^1_{n_1}, \ldots, u^m_1, \ldots, u^m_{n_m}) \in \mathbb{R}^n$ such that $\mu(\tilde{x}) := \sum_{i=1}^{m} \sum_{j=1}^{n_i} u^i_j x^i_j$ satisfies $\mu(\tilde{x}) \le \mu(\tilde{z})$ for any $\tilde{x} \in \bar{U}$. In particular, we have

$$\mu(\tilde{x}) \le \mu(\tilde{z}), \quad \forall \tilde{x} \in U. \tag{C3}$$

For each $i = 1, \ldots, m$, let $\phi_i : \{1, 2, \ldots, n_i\} \to \{1, 2, \ldots, n_i\}$ be a bijection such that $u^i_{\phi_i(1)} \le u^i_{\phi_i(2)} \le \cdots \le u^i_{\phi_i(n_i)}$, 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:

$$w^i_j := u^i_{\phi_i(j)}, \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, m, \tag{C4}$$

$$\tilde{w} := \left(w^1_1, \ldots, w^1_{n_1}, \ldots, w^m_1, \ldots, w^m_{n_m}\right), \tag{C5}$$

$$\lambda(\tilde{x}) := \sum_{i=1}^{m} \sum_{j=1}^{n_i} w^i_j x^i_j, \tag{C6}$$

then by definition $\tilde{w}$ satisfies condition (12). For any fixed $\tilde{x} \in U$, by Axiom C4, $\hat{\rho}((x^1_{\psi_1(1)}, \ldots, x^1_{\psi_1(n_1)}, \ldots, x^m_{\psi_m(1)}, \ldots, x^m_{\psi_m(n_m)})) = \hat{\rho}(\tilde{x}) < 1$, so $(x^1_{\psi_1(1)}, \ldots, x^1_{\psi_1(n_1)}, \ldots, x^m_{\psi_m(1)}, \ldots, x^m_{\psi_m(n_m)}) \in U$. Then, we have

$$\begin{aligned}\lambda(\tilde{x}) &= \sum_{i=1}^{m} \sum_{j=1}^{n_i} w^i_j x^i_j = \sum_{i=1}^{m} \sum_{j=1}^{n_i} u^i_{\phi_i(j)} x^i_j = \sum_{i=1}^{m} \sum_{j=1}^{n_i} u^i_{\phi_i(\psi_i(j))} x^i_{\psi_i(j)} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} u^i_j x^i_{\psi_i(j)} \\ &= \mu\left(x^1_{\psi_1(1)}, \ldots, x^1_{\psi_1(n_1)}, \ldots, x^m_{\psi_m(1)}, \ldots, x^m_{\psi_m(n_m)}\right) \le \mu(\tilde{z}) \quad (\text{by (C3)}).\end{aligned} \tag{C7}$$

Noting that $z^i_1 \le z^i_2 \le \cdots \le z^i_{n_i}$, $i = 1, 2, \ldots, m$, we obtain

$$\mu(\tilde{z}) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} u^i_j z^i_j \le \sum_{i=1}^{m} \sum_{j=1}^{n_i} u^i_{\phi_i(j)} z^i_j = \lambda(\tilde{z}). \tag{C8}$$

By (C7) and (C8), we have

$$\lambda(\tilde{x}) \le \lambda(\tilde{z}), \quad \forall \tilde{x} \in U. \tag{C9}$$

We shall show that the strict inequality holds in (C9). Suppose, by contradiction, that there exists $\tilde{r} \in U$ such that $\lambda(\tilde{r}) = \lambda(\tilde{z})$. With $w_{\max} := \max\{w^1_1, \ldots, w^1_{n_1}, \ldots, w^m_1, \ldots, w^m_{n_m}\}$, we have $\tilde{r} + \varepsilon\tilde{w} \le \tilde{r} + \varepsilon w_{\max}\mathbf{1}$ for any $\varepsilon > 0$. Thus, Axioms C1 and C2 yield

$$\hat{\rho}(\tilde{r} + \varepsilon\tilde{w}) \le \hat{\rho}(\tilde{r} + \varepsilon w_{\max}\mathbf{1}) = \hat{\rho}(\tilde{r}) + \varepsilon w_{\max}, \quad \forall \varepsilon > 0. \tag{C10}$$

Since $\hat{\rho}(\tilde{r}) < 1$, we have by (C10) that for small enough $\varepsilon > 0$, $\hat{\rho}(\tilde{r} + \varepsilon\tilde{w}) < 1$. Hence, $\tilde{r} + \varepsilon\tilde{w} \in U$ and (C9) implies $\lambda(\tilde{r} + \varepsilon\tilde{w}) \le \lambda(\tilde{z})$. 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

$$\lambda(\tilde{x}) < \lambda(\tilde{z}), \quad \forall \tilde{x} \in U. \tag{C11}$$

Since $\hat{\rho}(0) = 0$, we have $0 \in U$. Letting $\tilde{x} = 0$ in (C11) yields $\lambda(\tilde{z}) > 0$, so we can re-scale $\tilde{w}$ such that $\lambda(\tilde{z}) = 1 = \hat{\rho}(\tilde{z})$. Thus, (C11) becomes $\lambda(\tilde{x}) < 1$ for any $\tilde{x}$ such that $\hat{\rho}(\tilde{x}) < 1$, from which (C2) holds.

Lemma C.2. *Let $\mathcal{B}$ be defined in (A1). For any fixed $\tilde{z} \in \mathcal{B}$ and any $\hat{\rho}$ satisfying Axioms C1–C2, E1, and C4, there exists a weight $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^m) \in \mathbb{R}^n$ satisfying (10), (11), and (12), such that*

$$\hat{\rho}(\tilde{x}) \ge \sum_{i=1}^{m} \sum_{j=1}^{n_i} w^i_j x^i_j \quad \textit{for any } \tilde{x} \in \mathbb{R}^n, \qquad \textit{and} \qquad \hat{\rho}(\tilde{z}) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} w^i_j z^i_j. \tag{C12}$$

Proof. We will show this by considering two cases.

Case 1. $\hat{\rho}(\tilde{z}) = 1$. From Lemma C.1, there exists a weight $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^m) \in \mathbb{R}^n$ satisfying (12) such that the linear functional $\lambda(\tilde{x}) := \sum_{i=1}^{m} \sum_{j=1}^{n_i} w^i_j x^i_j$ satisfies (C1) and (C2).

Firstly, we prove that $\tilde{w}$ satisfies (10), which is equivalent to $\lambda(\mathbf{1}) = 1$. To this end, first note that for any $c < 1$, (C1) implies $\hat{\rho}(c\mathbf{1}) = c < 1$. Thus, (C2) implies $\lambda(c\mathbf{1}) < 1$, and, by continuity of $\lambda(\cdot)$, we obtain that $\lambda(\mathbf{1}) \le 1$. On the other hand, for any $c > 1$, (C1) implies $\hat{\rho}(2\tilde{z} - c\mathbf{1}) = 2\hat{\rho}(\tilde{z}) - c = 2 - c < 1$. Then it follows from (C1) and (C2) 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 $\lambda(\mathbf{1}) \ge 1$, and $\tilde{w}$ satisfies (10). Secondly, we prove that $\tilde{w}$ satisfies (11). For any fixed i and $1 \le j \le n_i$, let $k = n_1 + n_2 + \cdots + n_{i-1} + j$ and $\tilde{e} = (0, \ldots, 0, 1, 0, \ldots, 0)$ be the kth standard basis of $\mathbb{R}^n$. Then $w^i_j = \lambda(\tilde{e})$. For any $\varepsilon > 0$, Axioms C1 and C2 imply $\hat{\rho}(\tilde{z} - \tilde{e} - \varepsilon\mathbf{1}) = \hat{\rho}(\tilde{z} - \tilde{e}) - \varepsilon \le \hat{\rho}(\tilde{z}) - \varepsilon = 1 - \varepsilon < 1$. Then (C1) and (C2) 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, $w^i_j = \lambda(\tilde{e}) > -\varepsilon$, and the conclusion follows by letting ε go to 0. Thirdly, we prove that $\tilde{w}$ satisfies (C12). It follows from Axiom C1 and (C2) that

$$\forall c > 0, \ \ \lambda(\tilde{x}) < c \quad \text{for any } \tilde{x} \text{ such that } \hat{\rho}(\tilde{x}) < c. \tag{C13}$$

For any $c \le 0$, we choose $b > 0$ such that $b + c > 0$. Then it follows from (C13) 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

$$\forall c \le 0,\ \lambda(\tilde{x}) < c \quad \text{for any } \tilde{x} \text{ such that } \hat{\rho}(\tilde{x}) < c. \tag{C14}$$

It follows from (C13) and (C14) that $\hat{\rho}(\tilde{x}) \ge \lambda(\tilde{x})$ for any $\tilde{x} \in \mathbb{R}^n$, which in combination with $\hat{\rho}(\tilde{z}) = 1 = \lambda(\tilde{z})$ completes the proof of (C12).

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.

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 (12), the risk statistic (13) satisfies subadditivity for any $\tilde{x}$ and $\tilde{y} \in \mathbb{R}^n$. Let $(p_{i,1}, p_{i,2}, \ldots, p_{i,n_i})$ be the permutation of $(1, \ldots, n_i)$ such that $(\tilde{x}^i + \tilde{y}^i)_{p_{i,1}} \le (\tilde{x}^i + \tilde{y}^i)_{p_{i,2}} \le \cdots \le (\tilde{x}^i + \tilde{y}^i)_{p_{i,n_i}}$. Then for $k = 1, \ldots, n_i - 1$, the partial sum up to k satisfies

$$\sum_{j=1}^{k} (\tilde{x}^i + \tilde{y}^i)_{(j)} = \sum_{j=1}^{k} \big(\tilde{x}^i + \tilde{y}^i\big)_{p_{i,j}} = \sum_{j=1}^{k} \big(x^i_{p_{i,j}} + y^i_{p_{i,j}}\big) \ge \sum_{j=1}^{k} \big(x^i_{(j)} + y^i_{(j)}\big). \tag{C15}$$

In addition, we have for the total sum

$$\sum_{j=1}^{n_i} \big(\tilde{x}^i + \tilde{y}^i\big)_{(j)} = \sum_{j=1}^{n_i} \big(\tilde{x}^i + \tilde{y}^i\big)_{j} = \sum_{j=1}^{n_i} \big(x^i_j + y^i_j\big) = \sum_{j=1}^{n_i} \big(x^i_{(j)} + y^i_{(j)}\big). \tag{C16}$$

Rearranging the summation terms yields

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

which, along with (C15) and (C16), and because $w^i_j - w^i_{j+1} \le 0$, shows that

$$\begin{aligned}
\hat{\rho}(\tilde{x} + \tilde{y}) &\le \sup_{\tilde{w} \in \mathscr{W}} \left\{ \sum_{i=1}^{m} \left[\sum_{j=1}^{n_i - 1} (w^i_j - w^i_{j+1}) \sum_{k=1}^{j} (x^i_{(k)} + y^i_{(k)}) + w^i_{n_i} \sum_{k=1}^{n_i} (x^i_{(k)} + y^i_{(k)}) \right] \right\} \\
&= \sup_{\tilde{w} \in \mathscr{W}} \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n_i} w^i_j x^i_{(j)} + \sum_{i=1}^{m} \sum_{j=1}^{n_i} w^i_j y^i_{(j)} \right\} \\
&\le \sup_{\tilde{w} \in \mathscr{W}} \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n_i} w^i_j x^i_{(j)} \right\} + \sup_{\tilde{w} \in \mathscr{W}} \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n_i} w^i_j y^i_{(j)} \right\} = \hat{\rho}(\tilde{x}) + \hat{\rho}(\tilde{y}).
\end{aligned}$$

Secondly, we prove part (ii). Let $\mathscr{B}$ be defined in (A1). By Axiom C4, we only need to show that there exists a set of weights $\mathscr{W} = \{\tilde{w}\} \subset \mathbb{R}^n$ such that each $\tilde{w} \in \mathscr{W}$ satisfies (10), (11), and (12), and

$$\hat{\rho}(\tilde{x}) = \sup_{\tilde{w} \in \mathscr{W}} \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n_i} w^i_j x^i_j \right\} \quad \text{for } \forall \tilde{x} \in \mathscr{B}.$$

By Lemma C.2, for any $\tilde{y} \in \mathscr{B}$, there exists a weight $\widetilde{w(\tilde{y})} = (w(\tilde{y})^1_1, \ldots, w(\tilde{y})^1_{n_1}; \ldots; w(\tilde{y})^m_1, \ldots, w(\tilde{y})^m_{n_m})$ satisfying (10), (11), and (12), such that

$$\hat{\rho}(\tilde{x}) \ge \sum_{i=1}^{m} \sum_{j=1}^{n_i} w(\tilde{y})^i_j x^i_j \quad \text{for any } \tilde{x} \in \mathbb{R}^n, \qquad \text{and} \qquad \hat{\rho}(\tilde{y}) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} w(\tilde{y})^i_j y^i_j. \tag{C17}$$

Define $\mathscr{W}$ as the collection of such weights; i.e., $\mathscr{W} := \{\widetilde{w(\tilde{y})} \mid \tilde{y} \in \mathscr{B}\}$. Then each $\tilde{w} \in \mathscr{W}$ satisfies (10), (11), and (12). From (C17), for any fixed $\tilde{x} \in \mathscr{B}$, we have $\hat{\rho}(\tilde{x}) \ge \sum_{i=1}^{m} \sum_{j=1}^{n_i} w(\tilde{y})^i_j x^i_j$ for $\forall \tilde{y} \in \mathscr{B}$, and $\hat{\rho}(\tilde{x}) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} w(\tilde{x})^i_j x^i_j$. Therefore,

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

which completes the proof.

Appendix D. Proof of Theorem 3.4.

Proof. We assume, without loss of generality, that $s = 1$ in Axiom 4. 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

$$\hat{\rho}(c\tilde{x}) = c\hat{\rho}(\tilde{x}), \quad \forall c \geq 0, \forall \tilde{x} \geq 0. \tag{D1}$$

By Axiom 3, we have $\hat{\rho}(0) = \hat{\rho}(0) + \hat{\rho}(0)$, so $\hat{\rho}(0) = 0$. Axiom 3 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 $\mathbb{Q}^+$,

$$\hat{\rho}(q\tilde{x}) = q\hat{\rho}(\tilde{x}), \quad \forall q \in \mathbb{Q}^+, \tilde{x} \in \mathbb{R}^n. \tag{D2}$$

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 $n \to \infty$. 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 2 and (D2) 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 $n \to \infty$, we obtain (D1).

Now we are ready to prove the "only if" part. Let $\tilde{e}_j := (0, \dots, 0, 1, 0, \dots, 0)$ be the jth standard basis of $\mathbb{R}^n$, and $\ell_1 := 0$, $\ell_i := \sum_{j=1}^{i-1} n_j$, $i = 2, \dots, m$. By Axioms 1 and 3,

$$\begin{aligned}
\hat{\rho}(\tilde{x}) &= \hat{\rho}\big((x^1_{(1)}, x^1_{(2)}, \dots, x^1_{(n_1)}, \dots, x^m_{(1)}, x^m_{(2)}, \dots, x^m_{(n_m)})\big) \\
&= \hat{\rho}\left(\sum_{i=1}^{m}(0, 0, \dots, 0, x^i_{(1)}, x^i_{(2)}, \dots, x^i_{(n_i)}, 0, 0, \dots, 0)\right) \\
&= \sum_{i=1}^{m}\hat{\rho}\big((0, 0, \dots, 0, x^i_{(1)}, x^i_{(2)}, \dots, x^i_{(n_i)}, 0, 0, \dots, 0)\big).
\end{aligned} \tag{D3}$$

Further, by Axiom 3,

$$\begin{aligned}
&\hat{\rho}\big((0, \dots, 0, x^i_{(1)}, x^i_{(2)}, \dots, x^i_{(n_i)}, 0, \dots, 0)\big) \\
&\quad = \hat{\rho}\big((0, \dots, 0, 0, x^i_{(2)} - x^i_{(1)}, \dots, x^i_{(n_i)} - x^i_{(1)}, 0, \dots, 0)\big) + \hat{\rho}\left(x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) \\
&\quad = \hat{\rho}\big((0, \dots, 0, 0, 0, x^i_{(3)} - x^i_{(2)}, \dots, x^i_{(n_i)} - x^i_{(2)}, 0, \dots, 0)\big) + \hat{\rho}\left((x^i_{(2)} - x^i_{(1)}) \sum_{j=\ell_i+2}^{\ell_i+n_i} \tilde{e}_j\right) \\
&\qquad + \hat{\rho}\left(x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) \\
&\quad = \cdots \\
&\quad = \hat{\rho}\left((x^i_{(n_i)} - x^i_{(n_i-1)}) \sum_{j=\ell_i+n_i}^{\ell_i+n_i} \tilde{e}_j\right) + \cdots + \hat{\rho}\left((x^i_{(2)} - x^i_{(1)}) \sum_{j=\ell_i+2}^{\ell_i+n_i} \tilde{e}_j\right) + \hat{\rho}\left(x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) \\
&\quad = \big(x^i_{(n_i)} - x^i_{(n_i-1)}\big)\hat{\rho}\left(\sum_{j=\ell_i+n_i}^{\ell_i+n_i} \tilde{e}_j\right) + \cdots + (x^i_{(2)} - x^i_{(1)})\hat{\rho}\left(\sum_{j=\ell_i+2}^{\ell_i+n_i} \tilde{e}_j\right) + \hat{\rho}\left(x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right),
\end{aligned} \tag{D4}$$

where the last equality follows from (D1). If $x^i_{(1)} \geq 0$, then by (D1) we have

$$\hat{\rho}\left(x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) = x^i_{(1)}\hat{\rho}\left(\sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right). \tag{D5}$$

If $x^i_{(1)} < 0$, then because $x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j$ and $-x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j$ are scenario-wise comonotonic, we have by Axiom 3 that $\hat{\rho}(x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j) + \hat{\rho}(-x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j) = \hat{\rho}(0) = 0$, which implies

$$\hat{\rho}\left(x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) = -\hat{\rho}\left(-x^i_{(1)} \sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right) = x^i_{(1)}\hat{\rho}\left(\sum_{j=\ell_i+1}^{\ell_i+n_i} \tilde{e}_j\right), \tag{D6}$$

where the last equality follows from (D1). Then by (D3), (D4), (D5), and (D6), we obtain

$$\hat{\rho}(\tilde{x}) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} w_j^i x_{(j)}^i, \qquad \text{where} \quad w_j^i := \hat{\rho}\left(\sum_{k=\ell_i+j}^{\ell_i+n_i} \tilde{e}_k\right) - \hat{\rho}\left(\sum_{k=\ell_i+j+1}^{\ell_i+n_i} \tilde{e}_k\right).$$

Because by Axiom 2 $w_j^i \geq 0$ and $\sum_{i=1}^{m}\sum_{j=1}^{n_i} w_j^i = \hat{\rho}(\mathbf{1}) = 1$, the proof is completed.

Appendix E. Proof of Theorem 4.1 and Theorem 4.2.

Proof of Theorem 4.1 Let $\tilde{x}^i = (x_1^i, \ldots, x_{n_i}^i) \in \mathbb{R}^{n_i}$ be the data subset that is used to calculate VaR_{t-i}, $i = 1, \ldots, 60$, and $\tilde{x}^{i+60} = (x_1^{i+60}, \ldots, x_{n_{i+60}}^{i+60}) \in \mathbb{R}^{n_{i+60}}$ be the data subset used to calculate sVaR_{t-i}, $i = 1, \ldots, 60$. In addition, define the 121th scenario $\tilde{x}^{121} := 0 \in \mathbb{R}$ and $n_{121} := 1$. Let $n := \sum_{i=1}^{121} n_i$. We will show that (16) and (17) are natural risk statistics defined on $\mathbb{R}^n$. Define $\tilde{w} = (\tilde{w}^1, \ldots, \tilde{w}^{121}) = (w_1^1, \ldots, w_{n_1}^1, \ldots, w_1^{121}, \ldots, w_{n_{121}}^{121}) \in \mathbb{R}^n$ such that $w_j^i := 1_{\{j=\lceil 0.99 n_i\rceil\}}$, $1 \leq j \leq n_i$, $i = 1, \ldots, 121$. Then we have

$$\mathrm{VaR}_{t-i} = \sum_{j=1}^{n_i} w_j^i x_{(j)}^i, \qquad \mathrm{sVaR}_{t-i} = \sum_{j=1}^{n_{i+60}} w_j^{i+60} x_{(j)}^{i+60}; \quad i = 1, \ldots, 60. \tag{E1}$$

By (E1), the Basel II risk measure (16) is equal to

$$k \cdot \max\left\{\sum_{i=1}^{121}\sum_{j=1}^{n_i} u_j^i x_{(j)}^i, \sum_{i=1}^{121}\sum_{j=1}^{n_i} v_j^i x_{(j)}^i\right\}, \tag{E2}$$

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

$$\tilde{u}^1 := \frac{1}{k}\tilde{w}^1; \qquad \tilde{u}^i := 0, \quad i = 2, \ldots, 120; \qquad \tilde{u}^{121} := \frac{k-1}{k}\tilde{w}^{121},$$
$$\tilde{v}^i := \frac{1}{60}\tilde{w}^i, \quad i = 1, \ldots, 60; \qquad \tilde{v}^i := 0, \quad i = 61, \ldots, 121.$$

Hence, by Theorem 3.1, (16) is a natural risk statistic that corresponds to $s = k$ in Axiom C1. Again, by (E1), the Basel III risk measure (17) is equal to

$$k \cdot \max\left\{\sum_{i=1}^{121}\sum_{j=1}^{n_i} u_j^i x_{(j)}^i, \ \sum_{i=1}^{121}\sum_{j=1}^{n_i} v_j^i x_{(j)}^i\right\} + \ell \cdot \max\left\{\sum_{i=1}^{121}\sum_{j=1}^{n_i} g_j^i x_{(j)}^i, \sum_{i=1}^{121}\sum_{j=1}^{n_i} h_j^i x_{(j)}^i\right\}, \tag{E3}$$

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

$$\tilde{g}^i := 0, \ \forall i \neq 61 \text{ and } i \neq 121; \qquad \tilde{g}^{61} := \frac{1}{\ell}\tilde{w}^{61}; \qquad \tilde{g}^{121} = \frac{\ell-1}{\ell}\tilde{w}^{121},$$
$$\tilde{h}^i := 0, \ i = 1, \ldots, 60; \qquad \tilde{h}^i := \frac{1}{60}\tilde{w}^i, \ i = 61, \ldots, 120; \qquad \tilde{h}^{121} = 0.$$

It is straightforward to verify that (E3) satisfies Axioms C1–C4, with $s = k + \ell$ in Axiom C1. Hence, (17) is also a natural risk statistic.

Proof of Theorem 4.2 The IRC risk measure (18) 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.

Appendix F. Analysis of the robustness of MS and ES. The following tools in robust statistics show that MS is more robust than ES is.

(i) The *influence function* is an important tool for assessing the robustness of statistics. Let F be the distribution function of X, $\tilde{x} = (x_1, \ldots, x_n)$ 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 $T(\cdot)\colon \mathcal{M} \to \mathbb{R}$. MS and ES are both such functionals, since

$$\mathrm{MS}_\alpha(F) = F^{-1}\left(\frac{1+\alpha}{2}\right), \qquad \mathrm{ES}_\alpha(F) = \frac{1}{1-\alpha}\int_\alpha^1 F^{-1}(s)\,ds, \tag{F1}$$

where the second equality follows in Tasche [47, Proposition 3.4]. A natural estimator for $T(F)$ is $T(F_n)$, and

$$\mathrm{MS}_\alpha(F_n) = x_{(\lceil n(1+\alpha)/2\rceil)}, \tag{F2}$$

$$\mathrm{ES}_\alpha(F_n) = \frac{k-n\alpha}{(1-\alpha)n}x_{(k)} + \frac{1}{(1-\alpha)n}\sum_{j=k}^{n-1} x_{(j+1)},\ k=\lceil n\alpha\rceil. \tag{F3}$$

The robustness of the statistic $T(F_n)$ can be asymptotically characterized by its influence function (IF) $\mathrm{IF}(y,T,F) := \lim_{\varepsilon\downarrow 0}(1/\varepsilon)[T((1-\varepsilon)F+\varepsilon\delta_y)-T(F)]$, $y\in\mathbb{R}$, where δ_y is the point mass 1 at y that represents a contamination point to the distribution F. If the influence function is bounded; i.e., $\sup_{y\in\mathbb{R}}|\mathrm{IF}(y,T,F)|<\infty$, 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 [26]). The following proposition shows that ES has an unbounded influence function but MS has a bounded one.

Proposition F.1. (1) *If F has a probability density $f(\cdot)$ that is continuous and positive at $\mathrm{MS}_\alpha(F)$, then the influence function of MS_α is given by*

$$\mathrm{IF}(y,\mathrm{MS}_\alpha,F) = \begin{cases} \frac{1}{2}(\alpha-1)/f(\mathrm{MS}_\alpha(F)), & \text{if } y<\mathrm{MS}_\alpha(F),\\ 0, & \text{if } y=\mathrm{MS}_\alpha(F),\\ \frac{1}{2}(1+\alpha)/f(\mathrm{MS}_\alpha(F)), & \text{if } y>\mathrm{MS}_\alpha(F). \end{cases} \tag{F4}$$

(2) *If F has a positive probability density $f(\cdot)$, then the influence function of ES_α is given by*

$$\mathrm{IF}(y,\mathrm{ES}_\alpha,F) = \begin{cases} F^{-1}(\alpha)-\mathrm{ES}_\alpha(F), & \text{if } y\le F^{-1}(\alpha),\\ \dfrac{y}{1-\alpha}-\mathrm{ES}_\alpha(F)-\dfrac{\alpha}{1-\alpha}F^{-1}(\alpha), & \text{if } y>F^{-1}(\alpha). \end{cases} \tag{F5}$$

Proof. Because $\mathrm{MS}_\alpha(F)=F^{-1}((1+\alpha)/2)$, (F4) follows in Staudte and Sheather [44, Equation (3.2.3)]. To show (F5), define $F_{\varepsilon,y}(z) := (1-\varepsilon)F(z)+\varepsilon\delta_y(z)$, $z\in\mathbb{R}$. Then by definition,

$$F_{\varepsilon,y}(z) = \begin{cases} (1-\varepsilon)F(z), & \text{if } z<y,\\ (1-\varepsilon)F(z)+\varepsilon, & \text{if } z\ge y. \end{cases}$$

It follows in Tasche [47, Definition 3.2] that

$$\mathrm{ES}_\alpha(F) = \frac{1}{1-\alpha}\int_{[F^{-1}(\alpha),\infty)} zF(dz) - \frac{\alpha}{1-\alpha}F^{-1}(\alpha) + \frac{1}{1-\alpha}F^{-1}(\alpha)F\big(F^{-1}(\alpha)-\big).$$

Then we have

$$\mathrm{ES}_\alpha(F_{\varepsilon,y}) = \frac{1}{1-\alpha}\int_{[F_{\varepsilon,y}^{-1}(\alpha),\infty)} zF_{\varepsilon,y}(dz) - \frac{\alpha}{1-\alpha}F_{\varepsilon,y}^{-1}(\alpha) + \frac{1}{1-\alpha}F_{\varepsilon,y}^{-1}(\alpha)F_{\varepsilon,y}\big(F_{\varepsilon,y}^{-1}(\alpha)-\big). \tag{F6}$$

To compute $\mathrm{IF}(y,\mathrm{ES}_\alpha,F)$, we need to consider three cases:

Case 1. $y<F^{-1}(\alpha)$. In this case, for $\varepsilon>0$ small enough, $F_{\varepsilon,y}^{-1}(\alpha)=F^{-1}((\alpha-\varepsilon)/(1-\varepsilon))$, and $F_{\varepsilon,y}(F_{\varepsilon,y}^{-1}(\alpha)-)=F_{\varepsilon,y}(F^{-1}((\alpha-\varepsilon)/(1-\varepsilon))-)=(1-\varepsilon)F(F^{-1}(\alpha-\varepsilon)/(1-\varepsilon)+\varepsilon=\alpha$. And then by (F6), for $\varepsilon>0$ small enough,

$$\begin{aligned} G(\varepsilon) := \mathrm{ES}_\alpha(F_{\varepsilon,y}) &= \frac{1}{1-\alpha}\int_{[F_{\varepsilon,y}^{-1}(\alpha),\infty)} zF_{\varepsilon,y}(dz)\\ &= \frac{1-\varepsilon}{1-\alpha}\int_{[F^{-1}((\alpha-\varepsilon)/(1-\varepsilon)),\infty)} zF(dz) + \frac{\varepsilon}{1-\alpha}y1_{\{y\ge F^{-1}((\alpha-\varepsilon)/(1-\varepsilon))\}} = \frac{1-\varepsilon}{1-\alpha}\int_{[F^{-1}((\alpha-\varepsilon)/(1-\varepsilon)),\infty)} zF(dz). \end{aligned}$$

Hence,

$$\begin{aligned} \mathrm{IF}(y,\mathrm{ES}_\alpha,F) = G'(0) &= -\frac{1}{1-\alpha}\int_{[F^{-1}((\alpha-\varepsilon)/(1-\varepsilon)),\infty)} zF(dz)\bigg|_{\varepsilon=0}\\ &\quad + \frac{1-\varepsilon}{1-\alpha}(-1)F^{-1}\left(\frac{\alpha-\varepsilon}{1-\varepsilon}\right)f\left(F^{-1}\left(\frac{\alpha-\varepsilon}{1-\varepsilon}\right)\right)\frac{d}{d\varepsilon}F^{-1}\left(\frac{\alpha-\varepsilon}{1-\varepsilon}\right)\bigg|_{\varepsilon=0}\\ &= -\frac{1}{1-\alpha}\int_{[F^{-1}(\alpha),\infty)} zF(dz) + F^{-1}(\alpha) \end{aligned} \tag{F7}$$

Case 2. $y = F^{-1}(\alpha)$. In this case, $F_{\varepsilon,y}^{-1}(\alpha) = F^{-1}(\alpha)$, and $F_{\varepsilon,y}(F_{\varepsilon,y}^{-1}(\alpha)-) = F_{\varepsilon,y}(F^{-1}(\alpha)-) = (1-\varepsilon)\cdot F(F^{-1}(\alpha)) = (1-\varepsilon)\alpha$. And by (F6),

$$\begin{aligned} G(\varepsilon) = \mathrm{ES}_\alpha(F_{\varepsilon,y}) &= \frac{1}{1-\alpha}\int_{[F^{-1}(\alpha),\infty)} zF_{\varepsilon,y}(dz) - \frac{\varepsilon\alpha}{1-\alpha}F^{-1}(\alpha) \\ &= \frac{1-\varepsilon}{1-\alpha}\int_{[F^{-1}(\alpha),\infty)} zF(dz) + \varepsilon F^{-1}(\alpha). \end{aligned}$$

Hence,

$$\mathrm{IF}(y, \mathrm{ES}_\alpha, F) = G'(0) = -\frac{1}{1-\alpha}\int_{[F^{-1}(\alpha),\infty)} zF(dz) + F^{-1}(\alpha). \tag{F8}$$

Case 3. $y > F^{-1}(\alpha)$. In this case, for $\varepsilon > 0$ small enough, $F_{\varepsilon,y}^{-1}(\alpha) = F^{-1}(\alpha/(1-\varepsilon))$, and $F_{\varepsilon,y}(F_{\varepsilon,y}^{-1}(\alpha)-) = F_{\varepsilon,y}(F^{-1}(\alpha/(1-\varepsilon))-) = (1-\varepsilon)F(F^{-1}(\alpha/(1-\varepsilon))) = \alpha$. And then by (F6), for $\varepsilon > 0$ small enough,

$$\begin{aligned} G(\varepsilon) = \mathrm{ES}_\alpha(F_{\varepsilon,y}) &= \frac{1}{1-\alpha}\int_{[F_{\varepsilon,y}^{-1}(\alpha),\infty)} zF_{\varepsilon,y}(dz) = \frac{1-\varepsilon}{1-\alpha}\int_{[F^{-1}(\alpha/(1-\varepsilon)),\infty)} zF(dz) + \frac{\varepsilon}{1-\alpha}y1_{\{y\ge F^{-1}(\alpha/(1-\varepsilon))\}} \\ &= \frac{1-\varepsilon}{1-\alpha}\int_{[F^{-1}(\alpha/(1-\varepsilon)),\infty)} zF(dz) + \frac{\varepsilon}{1-\alpha}y. \end{aligned}$$

Hence,

$$\begin{aligned} \mathrm{IF}(y, \mathrm{ES}_\alpha, F) = G'(0) &= \frac{y}{1-\alpha} - \frac{1}{1-\alpha}\int_{[F^{-1}(\alpha/(1-\varepsilon)),\infty)} zF(dz)\bigg|_{\varepsilon=0} \\ &\quad + \frac{1-\varepsilon}{1-\alpha}(-1)F^{-1}\left(\frac{\alpha}{1-\varepsilon}\right)f\left(F^{-1}\left(\frac{\alpha}{1-\varepsilon}\right)\right)\frac{d}{d\varepsilon}F^{-1}\left(\frac{\alpha}{1-\varepsilon}\right)\bigg|_{\varepsilon=0} \\ &= \frac{y}{1-\alpha} - \frac{1}{1-\alpha}\int_{[F^{-1}(\alpha),\infty)} zF(dz) - \frac{\alpha}{1-\alpha}F^{-1}(\alpha). \end{aligned} \tag{F9}$$

Then (F5) follows from (F7), (F8), and (F9).

(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 Huber and Ronchetti [26, §1.4] for the mathematical definition. Hence, a high breakdown point is clearly desirable. It follows from Huber and Ronchetti [26, Theorem 3.7] and Equation (F1) that the asymptotic breakdown point of MS_α is $1-\alpha$ and the asymptotic breakdown point of ES_α is 0, which clearly shows the robustness of MS.

(iii) The *finite sample breakdown point* (see Huber and Ronchetti [26, Chap. 11]) of $\mathrm{MS}_\alpha(F_n)$ 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 $\mathrm{ES}_\alpha(F_n)$ is $1/(n+1)$, which means one additional corrupted sample can cause arbitrarily large bias to ES_α.

Acknowledgments. The authors 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. They particularly thank two anonymous referees for their constructive comments that help to improve the paper. They 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, and 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 Hong Kong RGC Direct Allocation Grant (Project DAG12SC05-3) and a grant from School-Based-Initiatives of HKUST (Project SBI11SC03). The preliminary versions of the paper were entitled "What is a good risk measure: Bridging the gaps between data, coherent risk measures, and insurance risk measures" and "What is a good external risk measure: Bridging the gaps between robustness, subadditivity, and insurance risk measures."

References

[1] Acerbi C, Nordio C, Sirtori C (2001) Expected shortfall as a tool for financial risk management. Preprint, Abaxbank, Italy.
[2] Acerbi C, Tasche D (2002) On the coherence of expected shortfall. *J. Bank. Finance* 26(7):1487–1503.
[3] Adrian T, Brunnermeier MK (2008) CoVaR. Federal Reserve Bank of New York Staff Reports 348, Federal Reserve Bank, New York.
[4] Ahmed S, Filipović D, Svindland G (2008) A note on natural risk statistics. *Oper. Res. Lett.* 36(6):662–664.
[5] Artzner P, Delbaen F, Eber J-M, Heath D (1999) Coherent measures of risk. *Math. Finance* 9(3):203–228.
[6] Basel Committee on Banking Supervision (2006) International convergence of capital measurement and capital standards: A revised framework (comprehensive version). Report, Bank for International Settlements, Basel, Switzerland.

[7] Basel Committee on Banking Supervision (2009) Guidelines for computing capital for incremental risk in the trading Book. Report, Bank for International Settlements, Basel, Switzerland.
[8] Basel Committee on Banking Supervision (2011) Revisions to the Basel II market risk framework. Report, Bank for International Settlements, Basel, Switzerland.
[9] Cont R, Deguest R, Scandolo G (2010) Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance* 10(6): 593–606.
[10] Daníelsson J, Jorgensen BN, Samorodnitsky G, Sarma M, de Vries CG (2005) Subadditivity re-examined: The case for value-at-risk. Working paper, London School of Economics.
[11] Delbaen F (2002) Coherent risk measures on general probability spaces. Sandmann K, Schönbucher PJ, eds. *Advances in Finance and Stochastics–Essays in Honour of Dieter Sondermann* (Springer, New York), 1–37.
[12] Denneberg D (1994) *Non-Additive Measure and Integral* (Kluwer Academic Publishers, Boston).
[13] Dhaene J, Goovaerts MJ, Kaas R (2003) Economic capital allocation derived from risk measures. *N. Am. Actuar. J.* 7(2):44–59.
[14] Föllmer H, Schied A (2002) Convex measures of risk and trading constraints. *Finance Stoch.* 6(4):429-447.
[15] Frittelli M, Gianin ER (2002) Putting order in risk measures. *J. Bank. Finance* 26(7):1473–1486.
[16] Garcia R, Renault É, Tsafack G (2007) Proper conditioning for coherent VaR in portfolio management. *Management Sci.* 53(3): 483–494.
[17] Gilboa I, Schmeidler D (1989) Maxmin expected utility with non-unique prior. *J. Math. Econom.* 18(2):141–153.
[18] Glasserman P (2012) Risk horizon and rebalancing horizon in portfolio risk measurement. *Math. Finance* 22(2):215–249.
[19] Gordy MB (2003) A risk-factor model foundation for ratings-based bank capital rules. *J. Financial Intermed.* 12(3):199–232.
[20] Gordy MB, Howells B (2006) Procyclicality in Basel II: Can we treat the disease without killing the patient? *J. Financial Intermed.* 15(3):395–417.
[21] Hansen LP, Sargent TJ (2007) *Robustness* (Princeton University Press, Princeton, NJ).
[22] Hart HLA (1994) *The Concept of Law*, 2nd ed. (Clarendon Press, Oxford, UK).
[23] Heyde CC, Kou SG (2004) On the controversy over tailweight of distributions. *Oper. Res. Lett.* 32(5):399–408.
[24] Hong C-S, Herk LF (1996) Incremental risk aversion and diversification preference. *J. Econom. Theory* 70(1):180–200.
[25] Huber PJ (1981) *Robust Statistics* (John Wiley & Sons, New York).
[26] Huber PJ, Ronchetti EM (2009) *Robust Statistics*, 2nd ed. (John Wiley & Sons, Hoboken, NJ).
[27] Ibragimov R (2004) On the robustness of economic models to heavy-tailedness assumptions. Mimeo, Yale University, New Haven, CT.
[28] Ibragimov R (2009) Portfolio diversification and value at risk under thick-tailedness. *Quant. Finance* 9(5):565–580.
[29] Ibragimov R, Walden J (2007) The limits of diversification when losses may be large. *J. Bank. Finance* 31(8):2551–2569.
[30] Jaschke S, Küchler U (2001) Coherent risk measures and good deal bounds. *Finance Stoch.* 5(2):181–200.
[31] Jorion P (2007) *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed. (McGraw-Hill, Boston).
[32] Keppo J, Kofman L, Meng X (2010) Unintended consequences of the market risk requirement in banking regulation. *J. Econom. Dynam. Control* 34(10):2192–2214.
[33] Kijima M (1997) The generalized harmonic mean and a portfolio problem with dependent assets. *Theory and Decision* 43(1):71–87.
[34] Kusuoka S (2001) On law invariant coherent risk measures. *Adv. Math. Econom.* 3:83–95.
[35] Litterman R (2005) Hot spots and hedges. Lehmann BN, eds. *The Legacy of Fischer Black* (Oxford University Press, New York), 55–95.
[36] Maccheroni F, Marinacci M, Rustichini A (2006) Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica* 74(6):1447–1498.
[37] McMinn RD (1984) A general diversification theorem: A note. *J. Finance* 39(2):541–550.
[38] McNeil A, Frey R, Embrechts P (2005) *Quantitative Risk Management* (Princeton University Press, Princeton, NJ).
[39] Rockafellar, RT, Uryasev S (2002) Conditional Value-at-Risk for general loss distributions. *J. Bank. Finance* 26(7):1443–1471.
[40] Samuelson, PA (1967) General proof that diversification pays. *J. Financial Quant. Anal.* 2(1):1–13.
[41] Schmeidler D (1989) Subjective probability and expected utility without additivity. *Econometrica* 57(3):571–587.
[42] Song Y, Yan J-A (2006) The representations of two types of functionals on $L^\infty(\Omega, \mathcal{F})$ and $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. *Sci. China Ser. A: Math.* 49(10):1376–1382.
[43] Song Y, Yan J-A (2009) Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. *Insurance Math. Econom.* 45(3):459–465.
[44] Staudte RG, Sheather SJ (1990) *Robust Estimation and Testing* (John Wiley & Sons, New York).
[45] Staum J (2004) Fundamental theorems of asset pricing for good deal bounds. *Math. Finance* 14(2):141–161.
[46] Tasche D (1999) Risk contributions and performance measurement. Preprint, Technical University of Munich, Munich, Germany.
[47] Tasche D (2002) Expected shortfall and beyond. *J. Bank. Finance* 26(7):1519–1533.
[48] Transportation Research Board of the National Academies (2003) Design speed, operating speed, and posted speed practices. National Cooperative Highway Research Report 504, Transportation Research Board, Washington, DC.
[49] Wang SS, Young VR, Panjer HH (1997) Axiomatic characterization of insurance prices. *Insurance Math. Econom.* 21(2):173–183.
[50] Yaari ME (1987) The dual theory of choice under risk. *Econometrica* 55(1):95–115.