On the Measurement of Economic Tail Risk*

Steven Kou† Xianhua Peng‡

Abstract

This paper attempts to provide a decision-theoretic 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 for the Choquet expected utility and the statistical property of elicitability (i.e. there exists an objective function such that minimizing the expected objective function yields the risk measure) is median shortfall, which is the median of tail loss distribution. Elicitability is important for backtesting. 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: C10, C44, C53, D81, G17, G18, G28, K23

*We are grateful to the seminar and conference participants at National University of Singapore, Peking University, University of Waterloo, Econometric Society China Meeting 2014, Econometric Society Asian Meeting 2014, Econometric Society European Meeting 2014, INFORMS Annual Meeting 2013, INFORMS Annual Meeting 2014, Quantitative Methods in Finance Conference 2013, RiskMinds Asia Conference 2013, Risk and Regulation Workshop 2014, the 41st European Group of Risk and Insurance Economists Seminar, and the 8th World Congress of the Bachelier Finance Society 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.

†Risk Management Institute and Department of Mathematics, National University of Singapore, 21 Heng Mui Keng Terrace, I3 Building 04-03, Singapore 119613. Email: matsteve@nus.edu.sg.

‡Corresponding author. Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Email: maxhpeng@ust.hk. Tel: (852)23587431. Fax: (852)23581643.

1 Introduction

This paper attempts to provide a decision-theoretic 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; see, e.g., Duffie and Pan (1997, 2001) and Jorion (2007).

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 expectated 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."

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-theoretic 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 (i.e., a scoring rule, see Winkler and Jose (2011)); then, the forecasting objective function can 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) point out 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 elicitability of a risk measure is also related to the concept of "consistency" of a risk measure proposed by Davis (2013), who shows that VaR exhibits some inherent superiority over other risk measures.

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.

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 statistical robustness, which means that a small deviation in the model or small changes in the data only results in a small change in the risk measurement. 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), and Hansen and Sargent (2001, 2007), among others. We also incorporate multiple models in this paper; see Section 3. We complement these papers by studying the link between risk measures and statistical uncertainty via elicitability. As for the second part of the meaning of robustness, Cont, Deguest and Scandolo (2010) show that expected shortfall leads to a less robust risk measurement procedure than historical VaR; Kou, Peng and Heyde (2006, 2013) propose a set of axioms for robust external risk measures, which include VaR.

There has been a growing literature on capital requirements for banking regulation and robust risk measurement. Glasserman and Kang (2013) investigate the design of risk weights to align regulatory and private objectives in a mean-variance framework for portfolio selection. Glasserman and Xu (2014) develop a framework for quantifying the impact of model error and for measuring and minimizing risk in a way that is robust to model error. Keppo, Kofman and Meng (2010) show that the Basel II market risk requirements may have the unintended consequence of postponing banks' recapitalization and hence increasing banks' default probability. We complement their work by applying our theoretical results to the study on which risk measure may be more suitable for setting capital requirements in Basel Accords; see Section 4.

Important contribution to measurement of risk based on economic axioms includes Aumann and Serrano (2008), Foster and Hart (2009, 2013), and Hart (2011), which study risk measurement of gambles (i.e., random variables with positive mean and taking 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.

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 is devoted to relevant comments.

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 $\mathcal{X}$ be a set of random variables that include all bounded random variables, i.e., $\mathcal{X} \supset \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$, where $\mathcal{L}^\infty(\Omega, \mathcal{F}, P) := \{X \mid \text{there exists } M < \infty \text{ such that } |X| \leq M, \text{a.s. P}\}$. 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 variance, then $\mathcal{X}$ can be specified as $\mathcal{L}^2(\Omega, \mathcal{F}, P)$; if ρ is VaR, then $\mathcal{X}$ can be specified as the set of all random variables.

An important relation between two random variables is comonotonicity (Schmeidler (1986)): Two random variables X and Y are said to be comonotonic, if $(X(\omega_1)-X(\omega_2))(Y(\omega_1)-Y(\omega_2)) \geq 0$, $\forall \omega_1, \omega_2 \in \Omega$. Let X and Y be the loss of two portfolios, respectively. Suppose that there is a representative agent in the economy and he or she prefers the profit $-X$ to the profit $-Y$. If the agent is risk averse, then his or her preference may imply that $-X$ is less risky than $-Y$. 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 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 $X \leq Y$.

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: $\rho(X) = \rho(Y)$ if X and Y have the same distribution.

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

Axiom A1 corresponds to the comonotonic independence axiom for the Choquet expected utility risk preferences (Schmeidler (1989)). Axiom A2 is a minimum requirement for a reasonable risk measure. Axiom A3 with $s = 1$ is used in Schmeidler (1986); the constant s in Axiom A3 can be related to the "countercyclical indexing" risk measures 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, Crockett, Goodhart, Persaud and Shin (2009), and Adrian and Shin (2014). Axiom A4 is standard for a law invariant risk measure. Axiom A5 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; h may not be left or right continuous. 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 $h \circ P(A) := h(P(A))$, $\forall A \in \mathcal{F}$.

Proof. See Appendix A. □

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.[1] Note that in the case of random variables, the corollary 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.[2] 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

[1] 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.

[2] 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 by (2), $\rho(X) \geq \rho(Y)$.

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. For $\alpha \in (0,1]$, ρ in (2) is equal to VaR_α if $h(x) := 1_{\{x>1-\alpha\}}$; ρ in (2) is equal to VaR_0 if $h(x) := 1_{\{x=1\}}$. VaR is monotonic with respect to first-order stochastic dominance.

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

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

For $\alpha = 1$, ES of X at level α is defined as $\mathrm{ES}_1(X) := F_X^{-1}(1)$. 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 slight modification of the conditional loss distribution. For $\alpha \in [0,1)$, $\rho(X)$ in (2) is equal to $\mathrm{ES}_\alpha(X)$ if

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

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

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))[3]

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

[3] 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 that defined in Kou, Peng and Heyde (2013). In fact, the definition in the aforementioned second paper is the same as the "tail conditional median" proposed in Kou, Peng and Heyde (2006).

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 and considers both the size and likelihood of losses 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[4]

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

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

$$\rho_\Delta(X) := \int_{(0,1]} F_X^{-1}(u)d\Delta(u), \tag{3}$$

where Δ 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.[5] A special case of (3) is the spectral risk measure (Acerbi (2002)), defined as

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

where $\tilde{\Delta}$ is a probability measure on $[0,1]$; it corresponds to (3) with Δ being specified as $\frac{d\Delta(u)}{du} = \int_{[0,u)} \frac{1}{1-y} d\tilde{\Delta}(y)$ for $u \in (0,1)$ and $\Delta(\{1\}) = \tilde{\Delta}(\{1\})$. 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.

[4]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 $\mathrm{MS}_0(X) = F_X^{-1}(\frac{1}{2}) = \mathrm{VaR}_{\frac{1}{2}}(X)$.

[5]In fact, for any fixed $u \in (0,1]$, $F_X^{-1}(u) = \mathrm{VaR}_u(X)$ as a functional on $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$ is a special case of the risk measure (2). By the proof of Lemma 2.1, VaR_u satisfies monotonicity, positive homogeneity, and comonotonic additivity, which implies that ρ_Δ satisfies Axioms A1-A4 for any Δ. On $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$, ρ_Δ automatically 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 it can be shown that q_α^+ cannot be represented by (3). Indeed, suppose for the sake of contradiction that there exists a Δ such that $q_\alpha^+(X) = \rho_\Delta(X)$, $\forall X \in \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$. Let X_0 have a strictly positive density on its support. Then, $F_{X_0}^{-1}(u)$ is continuous and strictly increases on $(0,1]$. Let $c > 0$ be a constant. Define $X_1 = X_0 \cdot 1_{\{X_0 \leq F_{X_0}^{-1}(\alpha)\}} + (X_0 + c) \cdot 1_{\{X_0 > F_{X_0}^{-1}(\alpha)\}}$. It follows from $q_\alpha^+(X_1) - q_\alpha^+(X_0) = \rho_\Delta(X_1) - \rho_\Delta(X_0)$ that ${}_\Delta((\alpha,1]) = 1$, which in combination with the strict monotonicity of $F_{X_0}^{-1}(u)$ implies that $\rho_\Delta(X_0) = \int_{(\alpha,1]} F_{X_0}^{-1}(u)\Delta(du) > F_{X_0}^{-1}(\alpha) = q_\alpha^+(X_0)$. This contradicts to $\rho_\Delta(X_0) = q_\alpha^+(X_0)$.

If a risk measure ρ satisfies Axiom A4 (law invariance), then $\rho(X)$ only depends on F_X; hence, ρ induces a statistical functional that maps a distribution F_X to a real number $\rho(X)$. For simplicity of notation, we still denote the induced statistical functional as ρ. Namely, we will use $\rho(X)$ and $\rho(F_X)$ interchangeably in the sequel.

2.2 Elicitability

In practice, the measurement of 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 $\hat{F}_X$ for forecasting the unknown true value $\rho(F_X)$. As one may come up with different procedures to forecast $\rho(F_X)$, it is an important issue to evaluate which procedure provides a better forecast of $\rho(F_X)$.

The theory of elicitability provides a decision-theoretic foundation for effective evaluation of point forecasting procedures. Suppose one wants to forecast the realization of a random variable Y using a point x, without knowing the true distribution F_Y. The expected forecasting error is given 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., $S(x, y) = (x - y)^2$ or $S(x, y) = |x - y|$. The optimal point forecast corresponding to 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, the median functional is elicitable, as minimizing the expected forecasting error with $S(x, y) = |x-y|$ yields the median functional. If ρ is elicitable, then one can evaluate two point forecasting methods by comparing their respective expected forecasting error $ES(x, Y)$. As F_Y is unknown, the expected 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.

If a statistical functional ρ is not elicitable, then for any objective function S, the minimization of the expected 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 errors, no matter what objective function S is used.

The concept of elicitability dates back to the pioneering work of Savage (1971), Thomson (1979), and Osband (1985) and is further developed by Lambert, Pennock and Shoham (2008) and Gneiting (2011), who contends that "in issuing and evaluating point forecasts, it is essential that either the objective function (i.e., the function S) 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.

In the present paper, we are concerned with the measurement of risk, which is given by a single-valued statistical functional. Following Definition 2 in Gneiting (2011), where the elicitability for a set-valued statistical functional is defined, we define the elicitability for a single-valued statistical functional as follows.[6]

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*

$$\rho(F) = \min\left\{ x \mid x \in \arg\min_x \int S(x, y) dF(y) \right\}, \ \forall F \in \mathcal{P}. \tag{4}$$

In the definition, we do not impose any condition on the objective function S.

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-theoretic foundation of Choquet expected utility (i.e., satisfying Axioms A1-A5), because the mean apparently does not capture tail risk.

Theorem 2.1. *Let $\rho : \mathcal{X} \to \mathbb{R}$ be a risk measure that satisfies Axioms A1-A5 and $\mathcal{X} \supset \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$. Let $\mathcal{P} := \{F_X \mid X \in \mathcal{X}\}$. Then, $\rho(\cdot)$ (viewed as a statistical*

[6] In Definition 2.1, 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 (4) 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.

functional on $\mathcal{P}$) 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 B. □

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 discontinuities on $[0,1]$; in particular, the proof is not based on any assumption on left or right continuity of $h(\cdot)$. The outline of the proof is as follows. First, we show that a 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 four kinds of risk measures have convex level sets: (i) $c\text{VaR}_0 + (1-c)\text{VaR}_1$ for some constant $c \in [0,1]$; (ii) VaR_α, $\alpha \in (0,1)$, and, in particular, MS_α, $\alpha \in [0,1)$; (iii) $\rho = cq_\alpha^- + (1-c)q_\alpha^+$, where $\alpha \in (0,1)$ and $c \in [0,1)$ are constants, $q_\alpha^-(F) := \inf\{x \mid F(x) \geq \alpha\}$, and $q_\alpha^+(F) := \inf\{x \mid F(x) > \alpha\}$; (iv) the mean functional. Lastly, we examine the elicitability of the aforementioned four kinds of risk measures; in particular, we show that $\rho = cq_\alpha^- + (1-c)q_\alpha^+$ for $c \in [0,1)$ is not elicitable by extending the main proposition in Thomson (1979).

3 Extension to Incorporate Multiple Models

The previous section address the issue of model uncertainty from the perspective of elicitability. Following Gilboa and Schmeidler (1989) and Hansen and Sargent (2001, 2007), we further incorporate robustness by considering multiple models (scenarios). More precisely, we consider m probability measures P_i, $i = 1, \dots, m$ on the state space $(\Omega, \mathcal{F})$. 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 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 ith scenario, the measurement of risk is given by ρ_i that satisfy Axioms A1-A5. Then by Lemma 2.1, ρ_i can be represented by $\rho_i(X) =$

$\int X d(h_i \circ P_i)$, where h_i is a distortion function, $i = 1, \dots, m$. We then propose the following risk measure to incorporate multiple scenarios:

$$\rho(X) = f(\rho_1(X), \rho_2(X), \dots, \rho_m(X)), \tag{5}$$

where $f : \mathbb{R}^m \to \mathbb{R}$ is called 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 $x_i \leq y_i, i = 1, \dots, m$.

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 an affine function of that of X. Axiom B2 states that if the risk measurement of X is less than or equal to that of Y under each scenario, then the aggregate risk measurement of X is also less than or equal to that of Y. 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 robustness.

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 $\tilde{w} = (w_1, \dots, w_m) \in \mathcal{W}$ satisfying $w_i \geq 0$ and $\sum_{i=1}^m w_i = 1$, such that*

$$f(\tilde{x}) = s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^m w_i x_i \right\}, \forall \tilde{x} \in \mathbb{R}^m. \tag{6}$$

Proof. First, we show that Axioms B1-B3 are equivalent to the Axioms C1-C4 in Kou, Peng and Heyde (2013) with $n_i = 1$, $i = 1, \dots, m$. Axioms B1 and B2 are the same as the Axioms C1 and C2, respectively. Axiom C4 holds for any function when $n_i = 1$, $i = 1, \dots, m$. Axioms C1 and C3 apparently implies Axiom B3. We will then show that Axiom B1 and B3 imply 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 Axioms B1 and B3 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). $\square$

In the representation (6), each weight $\tilde{w} \in \mathcal{W}$ can be regarded as a prior probability on the set of 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:[7]

$$\rho(X) = s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^{m} w_i \int X \, d(h_i \circ P_i) \right\}. \tag{7}$$

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

$$\rho(X) = s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{i=1}^{m} w_i \mathrm{MS}_{i,\alpha_i}(X) \right\}, \tag{8}$$

where $\mathrm{MS}_{i,\alpha_i}(X)$ is the median shortfall of X at confidence level α_i calculated under the ith scenario (model). The risk measure ρ in (8) addresses the issue of model uncertainty and incorporate robustness from two aspects: (i) under each scenario i, MS_{i,α_i} is elicitable and statistically robust (Kou, Peng and Heyde (2006, 2013) and Cont, Deguest and Scandolo (2010)); (ii) ρ incorporates multiple scenarios and multiple priors on the set of scenarios.

4 Application to Basel Accord Capital Rule for Trading Books

What risk measure should be used for setting capital requirements 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 tth day is specified 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 ith model, $i = 1, \ldots, 60$. Define the 61th model under which $X = 0$ with probability one. Then,

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

the Basel II risk measure is a special case of the class of risk measures considered in (8); 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 procyclicality 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 (8).

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 tth 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 proposed Basel 3.5 risk measure is a special case of the class of risk measures considered in (7).[8]

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

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

MS may be preferable than ES for setting capital requirements in banking regulation because (i) MS is elicitable but ES is not; and (ii) MS is robust but ES is not (Kou, Peng and Heyde (2006, 2013) and Cont, Deguest and Scandolo (2010)). Kou, Peng and Heyde (2013) show that robustness is indispensable for external risk measures used for legal enforcement such as calculating capital requirements.

[8]The Basel II, Basel 2.5, and newly proposed risk measure (Basel 3.5) for the trading book are also 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.

To further compare the robustness of MS with ES, we carry out a simple empirical study on the measurement of 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

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

- 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&P 500 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 ith 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.

Table 1: The comparison of the forecasts of one-day MS and ES of a portfolio of S&P 500 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 ith model, respectively, $i = 1, 2$. It is clear 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., $\mathrm{MS}_{\alpha,2} - \mathrm{MS}_{\alpha,1}$).

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%

5 Comments

5.1 Criticism of Value-at-Risk

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." Below are some popular criticisms of VaR in the literature.

(i) The VaR at level α does not provide information regarding the size of the tail loss distribution beyond VaR_α. However, the median shortfall at level α does address this issue by measuring the median size of the tail loss distribution beyond VaR_α.

(ii) There is a pathological counterexample that, for some level α, the VaR_α of a fully concentrated portfolio might be smaller than that of a fully diversified portfolio, which is against the economic intuition that diversification reduces risk; see Example 6.7 in McNeil et al. (2005, p. 241). However, this counterexample disappears if $\alpha > 98\%$.

(iii) VaR does not satisfy the mathematical axiom of subadditivity (Huber (1981), Artzner et al. (1999))[9]. However, the subadditivity axiom is somewhat controversial:[10] (1) The subadditivity axiom 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. (2) 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. (3) Although subadditivity ensures

[9] 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.

[10] 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 satisfy global subadditivity. Ibragimov and Walden (2007) and Ibragimov (2009) show that VaR is subadditive for the infinite variance stable 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., Cauchy distribution)" (Gaglianone, Lima, Linton and Smith (2011)). Garcia, Renault and Tsafack (2007) stress that "tail thickness required [for VaR] to violate subadditivity, even for small probabilities, remains an extreme situation because it corresponds to such poor conditioning information that expected loss appears to be infinite."

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.[11] (4) 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.[12] (5) Subadditivity is not necessarily needed for capital allocation or asset allocation.[13] (6) It is often argued that if a non-subadditive risk measure is used in determining the regulatory capital for a financial institution, then to reduce its regulatory capital, the institution has an incentive to legally break up into various subsidiaries. However, breaking up an institution into subsidiaries may not be bad, as it prevents the loss of one single business unit from causing the bankruptcy of the whole institution. On the contrary, if a subadditive risk measure is used, then that institution has an incentive to merge with other financial institutions, which may lead to financial institutions that are too big to fail. Hence, it is not clear by using this type of argument alone whether a risk measure should be subadditive or not.

(iv) Embrechts et al. (2014) argue that "with respect to dependence uncertainty in aggregation, VaR is less robust compared to expected shortfall" because VaR is not aggregation-robust but expected shortfall is. However, their counterexample (i.e., their Example 2.1) only shows that VaR may not be aggregation-robust at the level α such that $F^{-1}(\cdot)$ is not continuous at α. There are only at most a countable number of such α; in fact, if F is a continuous distribution, then no such α exists. On the contrary, for any other α, VaR at level α is aggregation-robust, because VaR at level α is Hampel-robust and Hampel-robustness implies aggregation-robustness; note that by Corollary 3.7 of Cont, Deguest and Scandolo (2010) expected shortfall is not Hampel-robust.

(iv) Expected shortfall is more conservative than VaR because $\mathrm{ES}_\alpha > \mathrm{VaR}_\alpha$. This

[11] 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 $\widehat{\mathrm{ES}}_\alpha(X_1)$ and $\widehat{\mathrm{ES}}_\alpha(X_2)$, which are estimates of $\mathrm{ES}_\alpha(X_1)$ and $\mathrm{ES}_\alpha(X_2)$ respectively. $\widehat{\mathrm{ES}}_\alpha(X_1) + \widehat{\mathrm{ES}}_\alpha(X_2)$ cannot be used as an upper bound for $\mathrm{ES}_\alpha(X_1 + X_2)$ because it is possible that $\widehat{\mathrm{ES}}_\alpha(X_1) + \widehat{\mathrm{ES}}_\alpha(X_2) < \mathrm{ES}_\alpha(X_1) + \mathrm{ES}_\alpha(X_2)$.

[12] 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 $X_1+X_2 \leq 5$. 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 $\rho(X_1) + \rho(X_2) \approx 20$ as the margin for the portfolio $X_1 + X_2$; instead, people will directly compute $\rho(X_1 + X_2)$.

[13] Kou, Peng and Heyde (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 Shi and Werker (2012), Wen, Peng, Liu, Bai and Sun (2013), Xi, Coleman and Li (2013), and the references therein for asset allocation methods using VaR and Basel Accord risk measures.

argument is misleading because ES at level α should be compared with VaR at level $(1+\alpha)/2$ (i.e. MS at level α). ES_α may be smaller (i.e., less conservative) than MS_α, as mean may be smaller than median. For example, if the tail loss distribution is a Weibull distribution with a shape parameter lager than 3.44, then ES_α is smaller than MS_α (see, e.g., Von Hippel (2005)).

5.2 Other 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 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. Weber (2006) derives a characterization theorem for risk measures with convex acceptance set $\mathcal{N}$ and convex rejection set $\mathcal{N}^c$ under some topological conditions;[14] that characterization theorem cannot be applied in this paper because we do not make any assumption on the forecasting objective function $S(\cdot,\cdot)$ in the definition of elicitability and hence the topological conditions may not hold. For example, the results in Bellini and Bignozzi (2013), which rely on the characterization theorem in Weber (2006), make strong assumptions on the forecasting objective function $S(\cdot,\cdot)$,[15] requiring 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 also related to the statistical theory for the evaluation of probability forecasts (Lai, Shen and Gross (2011)).

The axioms in this paper are based on economic considerations. Other axioms

[14] The characterization theorem (Theorem 3.1 of Weber (2006)) requires two topological conditions on $\mathcal{N}$: (1) there exists $x \in \mathbb{R}$ with $\delta_x \in \mathcal{N}$ such that for $y \in \mathbb{R}$ and $\delta_y \in \mathcal{N}^c$, $(1-\alpha)\delta_x + \alpha\delta_y \in \mathcal{N}$ for sufficiently small $\alpha > 0$; (2) $\mathcal{N}$ is ψ-weakly closed for some gauge function $\psi : \mathbb{R} \to [1,\infty)$.

[15] These assumptions include three conditions in Definition 3.1 and two conditions in Theorem 4.2: (1) $S(x,y)$ is continuous in y; (2) for any $x \in [-\epsilon,\epsilon]$ with $\epsilon > 0$, $S(x,y) \leq \psi(y)$ for some gauge function ψ.

based on mathematical considerations include convexity (Föllmer and Schied (2002), Frittelli and Gianin (2002, 2005)), comonotonic subadditivity (Song and Yan (2006, 2009), Kou, Peng and Heyde (2006, 2013)), comonotonic convexity (Song and Yan (2006, 2009)). Dhaene, Vanduffel, Goovaerts, Kaas, Tang and Vyncke (2006) provides a survey on comonotonicity and risk measures.

A Proof of Lemma 2.1

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 $s = 1$ in Axiom A3).

The "only if" part. First, we show that (2) holds for any $X \in \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$. 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 $\nu(\Omega) = 1$. For $M \geq 1$, define $\mathcal{L}^M := \{X \mid |X| \leq M\}$. 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 $\rho(X) = \rho(X^{M_0})$ (by Axiom A4) and $\nu(X > x) = \nu(X^{M_0} > x)$, $\forall x$. 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

$$
\begin{aligned}
\rho(X) = \rho(X^{M_0}) &= \int_0^\infty \nu(X^{M_0} > x)dx + \int_{-\infty}^0 (\nu(X^{M_0} > x) - 1)dx \\
&= \int_0^\infty \nu(X > x)dx + \int_{-\infty}^0 (\nu(X > x) - 1)dx. \qquad (9)
\end{aligned}
$$

Let U be a uniform $U(0,1)$ random variable. Define the function h such that $h(0) = 0$, $h(1) = 1$, and $h(p) := \rho(1_{\{U \leq p\}})$, $\forall p \in (0,1)$. By Axiom A4, $h(\cdot)$ satisfies $\nu(A) = h(P(A))$ for all A. Therefore, by (9), (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

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

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 h is a distortion function and ρ is defined by (2). Define the set function $\nu(A) := h(P(A)), \forall A \in \mathcal{F}$. Then $\rho(X)$ is the Choquet integral of X with respect to ν. By definition of ρ and simple verification, ρ satisfies Axioms A2-A5. It follows from Denneberg (1994, Proposition 5.1) that ρ satisfies positive homogeneity and comonotonic additivity, which implies that ρ satisfies Axiom A1. $\square$

B Proof of Theorem 2.1

First, we give the following definition:[16]

Definition B.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 $\rho(\lambda F_1 + (1-\lambda)F_2) = \rho(F_1)$, $\forall \lambda \in (0,1)$.*

The following Lemma B.1 gives a necessary condition for a single-valued statistical functional to be elicitable. The lemma is a variant of Proposition 2.5 of Osband (1985), Lemma 1 of Lambert, Pennock and Shoham (2008), and Theorem 6 of Gneiting (2011), which concern set-valued statistical functionals.

Lemma B.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 (4) 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 $\int S(t',y)dF_\lambda(y) \le \int S(t,y)dF_\lambda(y)$, which implies that $\lambda \int S(t',y)dF_1(y) + (1-\lambda)\int S(t',y)dF_2(y) \le \lambda \int S(t,y)dF_1(y) + (1-\lambda)\int S(t,y)dF_2(y)$. However, by definition of t, $\int S(t,y)dF_i(y) \le \int S(t',y)dF_i(y), i = 1,2$. 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 $t = \min\{x \mid x \in \arg\min_x \int S(x,y)dF_i(y)\}$,

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

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 B.2. *Let $c \in [0,1]$ be a constant. If ρ is defined in (2) with $h(u) = 1 - c, \forall u \in (0,1)$, $h(0) = 0$, and $h(1) = 1$, then $\rho = c\,VaR_0 + (1-c)\,VaR_1$, where $VaR_0(F) := \inf\{x \mid F(x) > 0\}$ and $VaR_1(F) := \inf\{x \mid F(x) = 1\}$. In addition, ρ has convex level sets with respect to $\mathcal{P} = \{F \mid \rho(F)$ is well defined$\}$.*

Proof. If $\mathrm{VaR}_0(F) \geq 0$, then

$$\begin{aligned}
\rho(F) = &\int_{(0,\mathrm{VaR}_0(F))} h(1-F(x))\,dx + \int_{(\mathrm{VaR}_0(F),\mathrm{VaR}_1(F))} h(1-F(x))\,dx \\
&+ \int_{(\mathrm{VaR}_1(F),\infty)} h(1-F(x))\,dx \\
= &\ \mathrm{VaR}_0(F) + (1-c)(\mathrm{VaR}_1(F) - \mathrm{VaR}_0(F)) = c\mathrm{VaR}_0(F) + (1-c)\mathrm{VaR}_1(F).
\end{aligned}$$

If $\mathrm{VaR}_0(F) < 0$, similar calculation also leads to $\rho(F) = c\mathrm{VaR}_0(F) + (1-c)\mathrm{VaR}_1(F)$.

Suppose $t = \rho(F_1) = \rho(F_2)$. Denote $F_\lambda := \lambda F_1 + (1-\lambda)F_2$, $\lambda \in (0,1)$. There are three cases:

(i) $c = 0$. Then, $t = \mathrm{VaR}_1(F_1) = \mathrm{VaR}_1(F_2)$. By definition of VaR_1, $F_i(x) < 1$ for $x < t$ and $F_i(x) = 1$ for $x \geq t$. Hence, for any $\lambda \in (0,1)$, it holds that $F_\lambda(x) < 1$ for $x < t$ and $F_\lambda(x) = 1$ for $x \geq t$. Hence, $\rho(F_\lambda) = \mathrm{VaR}_1(F_\lambda) = t$.

(ii) $c \in (0,1)$. Without loss of generality, suppose $\mathrm{VaR}_0(F_1) \leq \mathrm{VaR}_0(F_2)$. Since $t = c\mathrm{VaR}_0(F_1) + (1-c)\mathrm{VaR}_1(F_1) = c\mathrm{VaR}_0(F_2) + (1-c)\mathrm{VaR}_1(F_2)$, $\mathrm{VaR}_1(F_1) \geq \mathrm{VaR}_1(F_2)$. Hence, for any $\lambda \in (0,1)$, $\mathrm{VaR}_0(F_\lambda) = \mathrm{VaR}_0(F_1)$ and $\mathrm{VaR}_1(F_\lambda) = \mathrm{VaR}_1(F_1)$. Hence, $\rho(F_\lambda) = t$.

(iii) $c = 1$. Then, $t = \mathrm{VaR}_0(F_1) = \mathrm{VaR}_0(F_2)$. By definition of VaR_0, $F_i(x) = 0$ for $x < t$ and $F_i(x) > 0$ for $x > t$. Hence, for any $\lambda \in (0,1)$, it holds that $F_\lambda(x) = 0$ for $x < t$ and $F_\lambda(x) > 0$ for $x > t$. Hence, $\rho(F_\lambda) = \mathrm{VaR}_0(F_\lambda) = t$. $\square$

Lemma B.3. *Let $\alpha \in (0,1)$ and $c \in [0,1]$. Let ρ be defined in (2) with h being defined as $h(x) := (1-c) \cdot 1_{\{x=1-\alpha\}} + 1_{\{x>1-\alpha\}}$. 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 $\mathcal{P} = \{F_X \mid X$ is a proper random variable$\}$.

Proof. Define $g(x) := 1 - h(1-x)$, $x \in [0,1]$. Then, $g(x) = c \cdot 1_{\{x=\alpha\}} + 1_{\{x>\alpha\}}$, and ρ can be represented as

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

Note that $F(x) = \alpha$ for $x \in [q_\alpha^-(F), q_\alpha^+(F))$. Consider three cases:

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

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

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

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

(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 of (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 three cases:

(i) $c = 0$. Then, $\rho = q_\alpha^+$. Denote $t = q_\alpha^+(F_1) = q_\alpha^+(F_2)$, 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 $q_\alpha^-(F_1) \geq q_\alpha^-(F_2)$. 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.

(iii) $c = 1$. Then, $\rho = q_\alpha^- = \mathrm{VaR}_\alpha$. Denote $t = q_\alpha^-(F_1) = q_\alpha^-(F_2)$, 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 $q_\alpha^-(F_\lambda) = t$, i.e., q_α^- has convex level sets with respect to $\mathcal{P}$. $\square$

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

Theorem B.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 four cases holds:*

(i) There exists $c \in [0,1]$, such that $\rho = c\,VaR_0 + (1-c)\,VaR_1$, where $VaR_0(F) := \inf\{x \mid F(x) > 0\}$ and $VaR_1(F) := \inf\{x \mid F(x) = 1\}$.

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

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

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

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

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

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

Proof of Theorem B.1. 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 $0 \le x_1 < x_2 < \cdots < x_n$, $p_i > 0$, $i = 1, \ldots, n$, 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 $q \in (0,1)$, $g(q) = 0$. Then $g(u) = 1_{\{u=1\}}$. By Lemma B.2 (with $c = 0$), $\rho = \mathrm{VaR}_1$ and 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, $g(u) = 1_{\{u>0\}}$. By Lemma B.2 (with $c = 1$), $\rho = \mathrm{VaR}_0$ and ρ has convex level sets with respect to $\mathcal{P}$. (ii.ii) $\alpha \in (0,1)$ and $g(\alpha) = 1$. Then, $g(u) = 1_{\{u\ge\alpha\}}$. By Lemma B.3 (with $c = 1$), $\rho = q_\alpha^- = \mathrm{VaR}_\alpha$ and ρ has convex level sets with respect to $\mathcal{P}$. (ii.iii) $\alpha \in (0,1)$ and $g(\alpha) = 0$. Then, $g(u) = 1_{\{u>\alpha\}}$. By Lemma B.3 (with $c = 0$), $\rho = q_\alpha^+$ and 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 $(x_1, x_2) = (1 - c, -\frac{g(q)}{1-g(q)}(1-c) + \frac{1}{1-g(q)})$, $\forall c \in (0,1)$. 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 $\alpha \leq q_0 < 1$, $g(\alpha) \leq g(q_0) < 1$, $\beta \geq q_0 > 0$, and $g(\beta) \geq g(q_0) > 0$.

There are four subcases:

Case (iii.i) $\alpha = \beta$ 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 B.3, $\rho = cq_\alpha^- + (1-c)q_\alpha^+$ and ρ has convex level sets with respect to $\mathcal{P}$.

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, $g(\alpha+\epsilon) \geq g(\alpha) > 0$ 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 $g(\alpha+) \leq g((\alpha+\beta)/2) < 1$. Therefore, this case does not hold.

Case (iii.iii) $\alpha < \beta$, $g(\alpha) = 0$, 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, $g(\beta-) \geq g((\beta+\alpha)/2) > 0$. Hence, there exists $\epsilon_0 > 0$ such that $g(\beta-\epsilon) > 0$ for any $\epsilon \in (0,\epsilon_0)$. On the other hand, $g(\beta-\epsilon) \leq g(\beta) < 1$ 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 $g(\beta-) > 0$. Therefore, this case does not hold.

Case (iii.iv) $\alpha < \beta$, $g(\alpha) = 0$, $g(\beta) = 1$. Let $q_0 \in (0,1)$ such that $g(q_0) \in (0,1)$. Then, $\alpha < q_0 < \beta$. We will show that either there exists a constant $c \in (0,1)$ such that $g(u) = c$, $\forall u \in (0,1)$, or $g(u) = u$, $\forall u \in (0,1)$.

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 $g(\alpha+) \leq g(q_0) < 1$. 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

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

For any $q \in (0, 1)$ and $v \in (0, 1)$, $vq + 1 - v > q$ and $\lim_{v\uparrow 1}(vq + 1 - v) = q$. It then follows from (17) that

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

Second, we consider two cases for g:

Case (iii.iv.i) There exist $0 < u_1 < u_2 < 1$ such that $g(u_1) = g(u_2)$. Let $w_1 = \inf\{u \mid g(u) = g(u_1)\}$ and $w_2 = \sup\{u \mid g(u) = g(u_2)\}$. Consider three further cases: (a) $w_1 > 0$. Since $\lim_{q\downarrow w_1} \frac{1-u_2}{1-q} = \frac{1-u_2}{1-w_1} < 1 = \lim_{q\downarrow w_1} \frac{w_1}{q}$, there exists $q_0 \in (w_1, u_2)$ such that $\frac{1-u_2}{1-q_0} < \frac{w_1}{q_0}$. Choose $v_0 \in (0, 1)$ such that $\frac{1-u_2}{1-q_0} < v_0 < \frac{w_1}{q_0}$. Since $v_0 q_0 < w_1$, $g(v_0 q_0) < g(u_1)$. And, since $w_1 < q_0 < v_0 q_0 + 1 - v_0 < u_2$, $g(q_0) = g(v_0 q_0 + 1 - v_0) = g(u_1)$. Therefore, $-g(v_0 q_0) + \frac{g(q_0)}{1-g(q_0)}(1 - g(v_0 q_0 + 1 - v_0)) > 0$, which contradicts to (17). Hence, this case cannot hold. (b) $w_2 < 1$. Since $\lim_{q\uparrow w_2} \frac{1-w_2}{1-q} = 1 > \frac{u_1}{w_2} = \lim_{q\uparrow w_2} \frac{u_1}{q}$, there exists $q_0 \in (u_1, w_2)$ such that $\frac{1-w_2}{1-q_0} > \frac{u_1}{q_0}$. Choose $v_0 \in (0, 1)$ such that $\frac{1-w_2}{1-q_0} > v_0 > \frac{u_1}{q_0}$. Since $w_2 > q_0 > v_0 q_0 > u_1$, $g(q_0) = g(v_0 q_0) = g(u_1)$. And, since $v_0 q_0 + 1 - v_0 > w_2$, $g(v_0 q_0 + 1 - v_0) > g(u_1)$. Therefore, $-g(v_0 q_0) + \frac{g(q_0)}{1-g(q_0)}(1 - g(v_0 q_0 + 1 - v_0)) < 0$, which contradicts to (17). Hence, this case cannot hold. (c) $w_1 = 0$ and

$w_2 = 1$. In this case, $g(u) = c, \forall u \in (0,1)$, for some constant $c \in (0,1)$. By Lemma B.2, $\rho = c\mathrm{VaR}_0 + (1-c)\mathrm{VaR}_1$, and ρ has convex level sets with respect to $\mathcal{P}$.

Case (iii.iv.ii) g is strictly increasing on $(0,1)$. Then, $g(p_1) - g(p_2) \neq 0$ for any $p_1 \neq p_2$. 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

$$x_1 g(p_1) + x_2(1 - g(p_1)) = x_1 g(p_2) + (1 - g(p_2))x_3. \tag{19}$$

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

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

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

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

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

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

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

$$g(1-) = 1. \tag{24}$$

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

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

It follows from (25) and (24) that

$$g(0+) = \lim_{v \downarrow 0} g(\frac{v}{2}) = \lim_{v \downarrow 0} \frac{g(\frac{1}{2})}{1 - g(\frac{1}{2})}(1 - g(1 - \frac{v}{2})) = \frac{g(\frac{1}{2})}{1 - g(\frac{1}{2})}(1 - g(1-)) = 0. \tag{26}$$

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

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

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

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

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 $\{k/2^n, k = 1, \dots, 2^n - 1, n \in \mathbb{N}\}$ 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. $\square$

Finally, the proof of Theorem 2.1 is as follows.

Proof of Theorem 2.1. By Lemma B.1 and Theorem B.1, only those risk measures listed in cases (i)-(iv) of Theorem B.1 satisfy the necessary condition for being an elicitable risk measure. Therefore, we only need to study the elicitability of those risk measures.

First, we will show that for $c \in (0,1]$, $\rho = c\mathrm{VaR}_0 + (1-c)\mathrm{VaR}_1$ is not elicitable. Suppose for the sake of contradiction that ρ is elicitable, then there exists a function S such that (4) holds. For any u, letting $F = \delta_u$ in (4) and noting $\rho(\delta_u) = u$ 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 (4) yields $pS(cu + (1-c)v, u) + (1-p)S(cu + (1-c)v, v) \le pS(x,u) + (1-p)S(x,v)$, $\forall x$. Letting $p \to 0$ leads to

$$S(cu + (1-c)v, v) \le S(x,v), \forall u < v, \forall x. \tag{45}$$

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

$$S(cu + (1-c)v, v) \le S(v,v), \forall u < v. \tag{46}$$

By (44), $S(v,v) \le S(cu+(1-c)v,v)$, $\forall u < v$, which in combination with (46) implies $S(v,v) = S(cu+(1-c)v,v)$, $\forall u < v$; however, by (44), $S(v,v) = S(cu+(1-c)v,v)$ implies $v \le cu+(1-c)v$, which contradicts to $u < v$. Hence, ρ is not elicitable.

Second, we will show that for $c = 0$, $\rho = c\mathrm{VaR}_0 + (1-c)\mathrm{VaR}_1 = \mathrm{VaR}_1$ 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 $x < \rho(F)$,

$$\int_{\mathbb{R}} S(x,y)dF(y) = \int_{x<y\le\rho(F)} S(x,y)dF(y) = a\int_{x<y\le\rho(F)} dF(y) = a(1-F(x)) > 0.$$

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

Third, we will show that for any $\alpha \in (0,1)$, VaR_α is elicitable with respect to $\mathcal{P}$. Let $g(\cdot)$ be a strictly increasing function defined on $\mathbb{R}$. Define

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

and define $\mathcal{P} = \{F_X \mid E|g(X)| < \infty\}$.[17] It follows from Theorem 9 in Gneiting (2011) that

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

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 (4) with S defined in (47).

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

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

[17]For example, if $g(x) := x$, then $\mathcal{P} = \{F_X \mid X \in \mathcal{L}^1(\Omega,\mathcal{F},P)\}$; if $g(x) := x^{\frac{1}{2n+1}}$ $(n \ge 1)$, then $\mathcal{P}$ includes heavy tailed distributions with infinite mean such as Cauchy distribution.

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[18] there exist 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 for any $x \in [q_\alpha^-(F_0), q_\alpha^+(F_0)]$,

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

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

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

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

Since $c \in [0,1)$, $\rho(F_0) = cq_\alpha^-(F_0) + (1-c)q_\alpha^+(F_0) \in (q_\alpha^-(F_0), q_\alpha^+(F_0)]$. It then follows from (49), (50), (51), and (52) that for any $x \in [q_\alpha^-(F_0), q_\alpha^+(F_0)]$,

$$\begin{aligned}
&\int S(x,y)dF_0(y) - \int S(\rho(F_0),y)dF_0(y) \\
&= (A_1(x) - A_1(\rho(F_0)))\alpha + (A_2(x) - A_2(\rho(F_0)))(1-\alpha) = 0,
\end{aligned}$$

[18] Thomson (1979) obtains the proposition for the case when the interval $I = (-\infty, \infty)$; in our case, $I = (-a, a)$. It can be verified that the proof of 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 in combination with $\rho(F_0) \in \arg\min_x \int S(x,y)dF_0(y)$ implies that

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

Therefore,

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

which contradicts to $\rho(F_0) > q_\alpha^-(F_0)$. Hence, 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. □

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., Crockett, A., Goodhart, C., Persaud, A. D. and Shin, H. S. (2009). *The Fundamental Principles of Financial Regulation: 11th Geneva Report on the World Economy*, Centre for Economic Policy Research.

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

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.

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

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., Wang, B. and Wang, R. (2014). Aggregation-robustness and model uncertainty of regulatory risk measures, *Preprint*, ETH Zurich.

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. F. 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. (2013). A wealth-requirement axiomatization of riskiness, *Theoretical Economics* **8**(2): 591–620.

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

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

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 Statistics* **29**(1): 150–160.

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

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

Glasserman, P. and Kang, W. (2013). Design of risk weights, *Operations Research*, forthcoming.

Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk, *Quantitative Finance* **14**(1): 29–58.

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**(3): 395–417.

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.

Hart, S. (2011). Comparing risks by acceptance and rejection, *Journal of Political Economy* **119**(4): 617–638.

Huber, P. J. (1981). *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.

Keppo, J., Kofman, L. and Meng, X. (2010). Unintended consequences of the market risk requirement in banking regulation, *Journal of Economic Dynamics and Control* **34**(10): 2192 – 2214.

Kou, S., Peng, X. and Heyde, C. C. (2006). What is a good risk measure: Bridging the gaps between data, coherent risk measures, and insurance risk measures, *Preprint*, Columbia University.

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

Lai, T. L., Shen, D. and Gross, S. (2011). Evaluating probability forecasts, *Annals of Statistics* **39**(5): 2356–2382.

Lambert, N. S., Pennock, D. M. and Shoham, Y. (2008). Eliciting properties of probability distributions, *Proceedings of the 9th ACM Conference on Electronic Commerce*, EC '08, ACM, New York, NY, USA, pp. 129–138.

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.

McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techiniques, Tools*, Princeton University Press, Princeton, NJ.

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.

Shi, Z. and Werker, B. J. M. (2012). Short-horizon regulation for long-term investors, *Journal of Banking & Finance* **36**(12): 3227 – 3238.

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 $\mathcal{L}^\infty(\Omega, \mathcal{F})$ and $\mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P})$, *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**(3): 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.

Von Hippel, P. T. (2005). Mean, median, and skew: Correcting a textbook rule, *Journal of Statistical Education* **13**(2).

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

Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency, *Mathematical Finance* **16**(2): 419–442.

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

Winkler, R. L. and Jose, V. R. R. (2011). Scoring rules, *in* C. J. (ed.), *Wiley Encyclopedia of Operations Research and Management Science*, Vol. 7, John Wiley & Sons, New York, pp. 4733–4744.

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*, Academy of Mathematics and Systems Science, Chinese Academy of Sciences.

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