Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

On the Measurement of Economic Tail Risk

Steven Kou, Xianhua Peng

Operations Research
Vol. 64, No. 5, September–October 2016, pp. 1056–1072
ISSN 0030-364X (print) | ISSN 1526-5463 (online)
http://dx.doi.org/10.1287/opre.2016.1539

On the Measurement of Economic Tail Risk

Steven Kou
Risk Management Institute and Department of Mathematics, National University of Singapore, Singapore 119077, matsteve@nus.edu.sg

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

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 risk measures that satisfy a set of economic axioms for the Choquet expected utility and the statistical property of general elicitability (i.e., there exists an objective function such that minimizing the expected objective function yields the risk measure) are the mean functional and value-at-risk (VaR), in particular the median shortfall, which is the median of tail loss distribution and is also the VaR at a higher confidence level. We also discuss various approaches of backtesting and their relations to elicitability and co-elicitability; in particular, we show that the co-elicitability of VaR and expected shortfall does not lead to a reliable backtesting method for expected shortfall and there have been only indirect backtesting methods for expected shortfall. Furthermore, we 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; expected shortfall.
Subject classifications: financial institution: banks; forecasting: applications; government: regulations; probability: applications.
Area of review: Special Issue on Systemic Risk: Models and Mechanisms.
History: Received December 2014; revisions received April 2015, August 2015, October 2015, April 2016; accepted May 2016. Published online in *Articles in Advance* September 12, 2016.

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 et al. (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 violations 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, p. 19), "uncertainty can come from limited data, unknown models and misspecification of those models."

In the 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, p. 756) shows that α-quantile as a set-valued statistical functional 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 et al. (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, p. 791) 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 risk measures that satisfy both a set of economic axioms proposed by Schmeidler (1989) and the statistical requirement of general elicitability are the mean functional and value-at-risk, in particular the median shortfall, which is the median of the tail loss distribution and is also the VaR at a higher confidence level.

In this paper, we also provide comprehensive discussion on various approaches of backtesting and their relations to elicitability and the co-elicitability of more than one statistical functionals (Lambert et al. 2008), which is a *weaker* notion of elicitability than the notion of elicitability of one statistical functional. In particular, we show that (i) the co-elicitability of VaR and expected shortfall *does not* lead to a reliable backtesting method for expected shortfall, and (ii) there have been only *indirect* back-testing methods for expected shortfall; see Sections 2.4 and 2.5.

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 et al. (2006), and Hansen and Sargent (2001, 2007), among others. We also incorporate multiple models in this paper; see Section 3. We add to the literature by studying the link between risk measures and statistical uncertainty via elicitability. As for the second part of the meaning of robustness, Cont et al. (2010) show that expected shortfall leads to a less robust risk measurement procedure than historical VaR; Kou et al. (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 et al. (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 add to this literature 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 complements 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 and discusses various approaches of backtesting and their relations to elicitability and co-elicitability. 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 $\mathcal{L}^\infty(\Omega, \mathcal{F}, P) := \{X \mid \text{there exists } M < \infty \text{ such that } |X| \leqslant M, \text{ a.s. } P\}$. 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 $(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geqslant 0$, $\forall\, \omega_1, \omega_2 \in \Omega$. 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: $\rho(X) \leqslant \rho(Y)$, if $X \leqslant Y$.*

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 A1 is postulated based on two motivations. First, Axiom A1 is a very weak requirement on ρ, as the pairwise comonotonicity of three random variables is a very strong condition. Second, to see the intuition behind Axiom A1, consider three pairwise comonotonic random variables X, Y, Z and $\alpha \in (0,1)$. If $\rho(X) < \rho(Y)$, then it seems reasonable that $\rho(\alpha X) < \rho(\alpha Y)$. In addition, since X, Y, Z are comonotonic, adding $(1-\alpha)Z$ to αX or αY does not hedge away the risk of αX or αY; hence, it would be reasonable to have $\rho(\alpha X + (1-\alpha)Z) < \rho(\alpha Y + (1-\alpha)Z)$, which leads to Axiom A1. 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 et al. (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 need 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 1. *Let $(\Omega, \mathscr{F}, P)$ be a probability space on which a random variable with a uniform distribution on $[0,1]$ can be defined. Let $\mathscr{X} \supset \mathscr{L}^\infty(\Omega, \mathscr{F}, P)$ be a set of random variables ($\mathscr{X}$ may include unbounded random variables). A risk measure ρ: $\mathscr{X} \to \mathbb{R}$ satisfies Axioms A1–A5 if and only if there exists a distortion function $h(\cdot)$ such that*

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

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

where the integral in (1) is the Choquet integral of X with respect to the distorted nonadditive probability $h\circ P(A) := h(P(A))$, $\forall A \in \mathscr{F}$.

Proof. See E-Companion EC.1 (available as supplemental material at http://dx.doi.org/10.1287/opre.2016.1539). ☐

Lemma 1 extends the representation theorem in Wang et al. (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 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 is a quantile of the loss distribution at some predefined probability level. More precisely, let X be the random loss with general distribution function $F_X(\cdot)$, which may not be continuous or strictly increasing. For a given $\alpha \in (0,1]$, VaR of X at level α is defined as the left α-quantile of F:

$$\mathrm{VaR}_\alpha(X) := q_\alpha^-(F) := F_X^{-1}(\alpha) = \inf\{x \mid F_X(x) \geqslant \alpha\}. \tag{3}$$

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) with $s = 1$ is equal to VaR_α if $h(x) := 1_{\{x > 1-\alpha\}}$; ρ in (2) with $s = 1$ 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.,

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

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

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

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 $X \geqslant \mathrm{VaR}_\alpha(X)$; 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) with $s = 1$ is equal to $\mathrm{ES}_\alpha(X)$ if

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

For $\alpha = 1$, $\rho(X)$ in (2) with $s = 1$ 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 nonelicitability and nonrobustness. 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 et al. 2013)[3]

$$\begin{aligned} \mathrm{MS}_\alpha(X) &:= \text{median of the } \alpha\text{-tail distribution of } X \\ &= F_{\alpha,X}^{-1}\left(\tfrac{1}{2}\right) = \inf\left\{x \,\middle|\, F_{\alpha,X}(x) \geqslant \tfrac{1}{2}\right\}. \end{aligned}$$

For $\alpha = 1$, MS at level α is defined as $\mathrm{MS}_1(X) := F_X^{-1}(1)$. Therefore, MS at level α can *capture the tail risk 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]

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

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

Since $\mathrm{MS}_\alpha = \mathrm{VaR}_{(1+\alpha)/2}$, MS_α does not quantify the risk beyond $\mathrm{VaR}_{(1+\alpha)/2}$. However, it is also difficult to know the precise degree to which ES_α quantifies the risk beyond $\mathrm{VaR}_{(1+\alpha)/2}$; in fact, just as MS_α, ES_α can also fail to reveal large loss beyond $\mathrm{VaR}_{(1+\alpha)/2}$. For example, fix $c := \mathrm{VaR}_\alpha$ and consider a sequence of α-tail distributions $F_{\alpha,n}$ that are mixtures of translated exponential distributions and point mass distributions, which are defined by

$$F_{\alpha,n}(x) := \begin{cases} 0, & \text{for } x < c \\ (1-\beta(n))(1-e^{-\lambda(x-c)}) & \\ \qquad + \beta(n) 1_{\{n \leqslant x\}}, & \text{for } x \geqslant c, \end{cases} \tag{4}$$

where $\lambda, \mu > 0$, $\beta(n) := \mu/(n - c - (1/\lambda))$.

In other words, $F_{\alpha,n}$ is the mixture of $c + \exp(\lambda)$ (with probability $(1-\beta(n))$) and the point mass δ_n (with probability $\beta(n)$). Under $F_{\alpha,n}$, a large loss with size n occurs with a small probability $\beta(n)$. For each n, $\mathrm{ES}_{\alpha,n}$, which is the mean of $F_{\alpha,n}$, is always equal to $c + \mu + 1/\lambda$; hence, ES_α fails in the same way as MS_α regarding the detection of the large loss with size n, which may occur beyond $\mathrm{VaR}_{(1+\alpha)/2}$. This example shows that the degree to which ES_α quantifies the risk beyond $\mathrm{VaR}_{(1+\alpha)/2}$ might also be limited. After all, MS_α and ES_α are, respectively, the median and the mean of *the same* α-tail loss distribution. The information contained in the mean of a distribution might not be more than that contained in the median of the same distribution, and vice versa.

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

$$\rho_\Delta(X) := \int_{(0,1]} F_X^{-1}(u)\, d\Delta(u), \tag{5}$$

where Δ is a probability measure on $(0,1]$. The class of risk measures represented by (2) includes and are strictly larger than the class of generalized spectral risk measures, as they all satisfy Axioms A1–A5.[5] A special case of (5) is the spectral risk measure (Acerbi 2002, Definition 3.1), defined as

$$\rho(X) = \int_{(0,1)} F_X^{-1}(u) \phi(u)\, du, \tag{6}$$

where $\phi(\cdot)$ is increasing, nonnegative, and $\int_0^1 \phi(u)\, du = 1$.

Because of the requirement that ϕ is increasing, the class of spectral risk measure is much smaller than the class of generalized spectral risk measure defined in (5). The distinction between the spectral risk measure and that in (5) is that the former is convex but the latter may not be convex. The convexity requires that the function ϕ in (6) is an increasing function. 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 $h(x) = 1 - (1 - x^{1/(1+\alpha)})^{1+\alpha}$ in (2) with $s = 1$, where $\alpha \geqslant 0$ is a constant.

The class of risk measures satisfying Axioms A1–A5 and the class of law-invariant coherent (convex) risk measures have nonempty intersections but no one is the subset of the other. For example, expected shortfall belongs to both classes; VaR belongs to the former but not the latter. The class of risk measures satisfying Axioms A1–A5 include the class of law-invariant spectral risk measures as a strict subset. For example, VaR belongs to the former but not the latter. The class of risk measures satisfying Axioms A1–A5 is the same as the class of "distortion risk measure" proposed in Wang et al. (1997). The distortion risk measures sometimes refer to the class of risk measures defined in (5). As we point out in Example 4, the class of risk measures defined in (5) is a strict subset of the class of risk measures satisfying Axioms A1–A5.

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

The measurement of risk of X using ρ may be viewed as a point forecasting problem, because the risk measurement $\rho(X)$ (or $\rho(F_X)$) summarizes the distribution F_X by a real number $\rho(X)$, just as a point forecast for X does. In practice, the true distribution F_X is unknown and one has to estimate 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

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

where $S(x, y)\colon \mathbb{R}^2 \to \mathbb{R}$ 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

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

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 with respect to a specified class of distributions $\mathcal{P}$ if there exists a forecasting objective function S such that for any distribution $F \in \mathcal{P}$, minimizing the expected forecasting error yields $\rho(F)$. 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 $\mathrm{ES}(x, Y)$. As F_Y is unknown, the expected forecasting error can be approximated by the average $(1/n)\sum_{i=1}^n S(x_i, Y_i)$, where $Y_1, \ldots, Y_n$ are samples that have the distribution F_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 et al. (2008) and Gneiting (2011, p. 749), 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 et al. (2009) also points out the critical importance of the specification of an objective function or an elicitable target functional. In Gneiting (2011, Definition 2) defines the elicitability for a set-valued statistical functional T as follows.

Definition 1 (Definition 2 in Gneiting (2011)). A *set-valued* statistical functional T is elicitable with respect to a class of distributions $\mathcal{P}$ if there exists a forecasting objective function $S\colon \mathbb{R}^2 \to \mathbb{R}$ such that

$$T(F) = \arg\min_x \int S(x, y)\, dF(y), \quad \forall\, F \in \mathcal{P}. \tag{7}$$

In the statistics literature, the α-quantile is defined as a *set-valued* statistical functional that maps a distribution F to the set $\{x \mid \lim_{y\uparrow x} F(y) \leqslant \alpha \leqslant F(x)\} = [q_\alpha^-(F), q_\alpha^+(F)]$, where $q_\alpha^-(F) := \inf\{x \mid F(x) \geqslant \alpha\}$ and $q_\alpha^+(F) := \inf\{x \mid F(x) > \alpha\}$ are, respectively, the left α-quantile and the right α-quantile of F. It has been shown (see, e.g., Gneiting 2011, Theorem 9) that the α-quantile as a set-valued statistical functional is elicitable with respect to

$$\mathcal{D}^1 := \big\{F \mid F \text{ is a distribution on } \mathbb{R} \text{ and has finite first moment}\big\}, \tag{8}$$

and the corresponding forecasting objective function can be defined as

$$S_\alpha(x, y) = (1_{\{x \geqslant y\}} - \alpha)(x-y). \tag{9}$$

In the present paper, we are concerned with the measurement of risk, which is a *single-valued* statistical functional. In the finance literature (see, e.g., Artzner et al. 1999, Definition 3.3), the VaR at level α is defined as $\mathrm{VaR}_\alpha := q_\alpha^-$, i.e., the left α-quantile. As a single-valued statistical functional, $\mathrm{VaR}_\alpha = q_\alpha^-$ is elicitable with respect to $\mathcal{D}^1 \cap \{F \mid q_\alpha^-(F) = q_\alpha^+(F)\}$ but not elicitable with respect to $\mathcal{D}^1$, because for those F with $q_\alpha^-(F) < q_\alpha^+(F)$, $\arg\min_x \int S_\alpha(x, y)\, dF(y)$ is an interval $[q_\alpha^-(F), q_\alpha^+(F)]$ instead of a single point; this is a minor technical nuisance because for any given F, $\mathrm{VaR}_\alpha(F) = \arg\min_x \int S_\alpha(x, y)\, dF(y)$ holds for all $\alpha \in (0, 1)$ except for a countable set of α at which $q_\alpha^-(F) < q_\alpha^+(F)$. To avoid such a minor technical nuisance, we slightly generalize the definition of elicitability in Definition 1 to define the *general elicitability* for a single-valued statistical functional as follows.

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

$$\rho(F) = \min\left\{x \,\middle|\, x \in \arg\min_x \int S(x, y)\, dF(y)\right\}, \quad \forall\, F \in \mathcal{P}. \tag{10}$$

In the definition, we only require that S satisfies the condition that $\int S(x, y)\, dF(y)$ is well defined and finite for any $F \in \mathcal{P}$. We do not need other conditions such as continuity or smoothness on S. In Theorem 1 in Section 2.3, we will show that VaR_α is general elicitable with respect to $\mathcal{D}^1$, which is consistent with the fact that α-quantile as a set-valued statistical functional is elicitable (in the sense of Definition 1) with respect to $\mathcal{D}^1$ and thus eliminates the aforementioned minor technical nuisance.

We have the following simple lemma showing that the definition of the general elicitability coincides with and generalizes the definition of elicitability in Definition 1.

Lemma 2. *If a single-valued statistical functional $\rho(\cdot)$ is elicitable (in the sense of Definition 1) with respect to a class of distributions $\mathcal{P}$, then it is general elicitable with respect to $\mathcal{P}$.*

Proof. If $\rho(\cdot)$ is elicitable with respect to a class of distributions $\mathcal{P}$, then there exists a forecasting objective function S such that (7) holds, which implies that (10) holds. Therefore, $\rho(\cdot)$ is general elicitable with respect to $\mathcal{P}$. □

2.3. Main Result

Let $\mathcal{D}_{\mathrm{disc}}$ be the set of discrete distributions having positive probabilities only on a finite number of values. The following Theorem 1 shows that value at risk and the mean functional are the *only* risk measures that (i) are general elicitable with respect to $\mathcal{D}_{\mathrm{disc}}$; and (ii) have the decision-theoretic foundation of Choquet expected utility (i.e., satisfying Axioms A1–A5). In particular, value at risk at level $(1+\alpha)/2$, which is the median shortfall at level α, provides a precise description of the average size of loss beyond VaR_α by the median of the tail loss distribution; whereas the mean functional captures the tail risk in the sense that knowing $E(L)$ leads to an upper bound $(1/x)E(L)$ for the tail probability $P(L > x)$ if $L \geqslant 0$.[6]

Theorem 1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space on which a random variable with a uniform distribution on $[0,1]$ can be defined. Let $\rho\colon \mathcal{X} \to \mathbb{R}$ be a risk measure that satisfies Axioms A1–A5 and $\mathcal{X} \supset \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$. Let $\mathcal{P}^\rho := \{F_X \mid X \in \mathcal{X}\}$ and let $\mathcal{D}_{\mathrm{disc}}$ be the set of discrete distributions that have positive probabilities only on a finite number of values. Then, $\rho(\cdot)$ (viewed as a single-valued statistical functional on $\mathcal{P}^\rho$) is general elicitable with respect to $\mathcal{D}_{\mathrm{disc}}$ if and only if one of the following two cases holds:*

(i) *$\rho = \mathrm{VaR}_\alpha$ for some $\alpha \in (0,1]$ (noting that $\mathrm{MS}_\alpha = \mathrm{VaR}_{(\alpha+1)/2}$ for $\alpha \in [0,1]$). Here VaR_α is a single valued functional defined as $\mathrm{VaR}_\alpha(F) := q_\alpha^-(F) = \inf\{x \mid F(x) \geqslant \alpha\}$.*

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

Furthermore, VaR_1 is general elicitable with respect to $\mathcal{D}^\infty := \{F_X \mid X \in \mathcal{L}^\infty(\Omega, \mathcal{F}, P)\}$, and for any $\alpha \in (0,1)$, VaR_α and the mean functional $\int x\, dF(x)$ are general elicitable with respect to a larger class $\mathcal{D}^1$ defined in (8).

Proof. See E-Companion EC.2. □

Remark 1. Since $\mathcal{D}_{\mathrm{disc}} \subset \mathcal{D}^\infty$, general elicitability with respect to $\mathcal{D}^\infty$ implies general elicitability with respect to $\mathcal{D}_{\mathrm{disc}}$; in addition, both risk measures in case (i) and case (ii) are general elicitable with respect to $\mathcal{D}^\infty$. Hence, Theorem 1 will also hold if $\mathcal{D}_{\mathrm{disc}}$ in the statement of Theorem 1 is replaced by $\mathcal{D}^\infty$.

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 general elicitable with respect to $\mathcal{P}$ is that ρ has convex level sets with respect to $\mathcal{P}$, i.e., $\rho(F_1) = \rho(F_2)$ and $\lambda F_1 + (1-\lambda)F_2 \in \mathcal{P}$ imply that $\rho(F_1) = \rho(\lambda F_1 + (1-\lambda)F_2)$, $\forall \lambda \in (0,1)$, $\forall F_1, F_2 \in \mathcal{P}$. The second and the key step is to prove the following theorem.

Theorem 2. *Let $\mathcal{D}_{\mathrm{disc}}$ be the class of discrete distributions that have positive probabilities only on a finite number of values. Let h be a distortion function defined on $[0,1]$ and let $\rho(\cdot)$ be defined as in (2) with $s = 1$. Then, $\rho(\cdot)$ has convex level sets with respect to $\mathcal{D}_{\mathrm{disc}}$ if and only if one of the following four cases holds:*

(i) *There exists $c \in [0,1]$, such that $\rho = c\,\mathrm{VaR}_0 + (1-c)\,\mathrm{VaR}_1$, where $\mathrm{VaR}_0(F) := \inf\{x \mid F(x) > 0\}$ and $\mathrm{VaR}_1(F) := \inf\{x \mid F(x) = 1\}$ (i.e., $h(u) = 1 - c$, $\forall u \in (0,1)$).*

(ii) *There exists $\alpha \in (0,1)$ such that $\rho(F) = \mathrm{VaR}_\alpha(F)$, $\forall F$ (i.e., $h(u) = 1_{\{u > 1-\alpha\}}$).*

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

$$\rho(F) = c q_\alpha^-(F) + (1-c) q_\alpha^+(F), \quad \forall F, \tag{11}$$

where $q_\alpha^-(F) := \inf\{x \mid F(x) \geqslant \alpha\}$ and $q_\alpha^+(F) := \inf\{x \mid F(x) > \alpha\}$ (i.e., $h(u) = (1-c)\cdot 1_{\{u=1-\alpha\}} + 1_{\{u>1-\alpha\}}$).

(iv) *$\rho(F) = \int x\, dF(x)$, $\forall F$ (i.e., $h(u) = u$).*

Furthermore, the risk measures ρ listed in cases (i)–(iv) have convex level sets with respect to $\mathcal{P}^\rho \supseteq \mathcal{D}^\infty$ ($\mathcal{P}^\rho$ is defined in Theorem 1 and $\mathcal{D}^\infty := \{F_X \mid X \in \mathcal{L}^\infty(\Omega, \mathcal{F}, P)\}$).

Proof. See E-Companion EC.2. □

Lastly, we examine the general elicitability of the four kinds of risk measures in Theorem 2 with respect to $\mathcal{D}_{\mathrm{disc}}$; in particular, we show that $\rho = c\,\mathrm{VaR}_0 + (1-c)\,\mathrm{VaR}_1$ for $c \in (0,1]$ and $\rho = c q_\alpha^- + (1-c) q_\alpha^+$ for $c \in [0,1)$ are not general elicitable with respect to $\mathcal{D}_{\mathrm{disc}}$.

Remark 2. After reading the proof of Theorem 2 (i.e., Theorem A.1 in Kou and Peng 2014), Wang and Ziegel (2015) attempted to provide an alternative proof; however, that proof is incomplete as it missed the case (i) in the theorem.

The following Theorem 3, which identifies elicitable risk measures, is the counterpart of Theorem 1, which identifies general elicitable risk measures. The conclusion and proof of the two theorems are almost identical, which clearly shows that the general elicitability is just a slight generalization of elicitability to avoid minor technical nuisances.

THEOREM 3. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space on which a random variable with a uniform distribution on* $[0, 1]$ *can be defined. Let* $\rho\colon \mathcal{X} \to \mathbb{R}$ *be a risk measure that satisfies Axioms A1–A5 and* $\mathcal{X} \supset \mathcal{L}^\infty(\Omega, \mathcal{F}, P)$. *Let* $\mathcal{P}^\rho := \{F_X \mid X \in \mathcal{X}\}$ *and let* $\mathcal{D}_{\text{disc}}$ *be the set of discrete distributions that have positive probabilities only on a finite number of values. Let* $\alpha_0 \in (0, 1)$ *be fixed. Then,* $\rho(\cdot)$ (*viewed as a single-valued statistical functional on* $\mathcal{P}^\rho$) *is elicitable* (*in the sense of Definition 1*) *with respect to* $\mathcal{D}_{\text{disc}} \cap \{F \mid q^-_{\alpha_0}(F) = q^+_{\alpha_0}(F)\}$ *if and only if one of the following two cases holds*:

(i) $\rho = \text{VaR}_{\alpha_0} = q^-_{\alpha_0}$.

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

Furthermore, VaR_{α_0} *and the mean functional are elicitable* (*in the sense of Definition 1*) *with respect to a larger class* $\mathcal{D}^1 \cap \{F \mid q^-_{\alpha_0}(F) = q^+_{\alpha_0}(F)\}$.

PROOF. See E-Companion EC.3. □

REMARK 3. Since $\mathcal{D}_{\text{disc}} \subset \mathcal{D}^\infty$, elicitability with respect to $\mathcal{D}^\infty$ implies elicitability with respect to $\mathcal{D}_{\text{disc}}$; in addition, both VaR_{α_0} and the mean functional are elicitable with respect to $\mathcal{D}^1 \cap \{F \mid q^-_{\alpha_0}(F) = q^+_{\alpha_0}(F)\} \supset \mathcal{D}^\infty \cap \{F \mid q^-_{\alpha_0}(F) = q^+_{\alpha_0}(F)\}$. Hence, Theorem 3 will also hold if $\mathcal{D}_{\text{disc}}$ in the statement of Theorem 3 is replaced by $\mathcal{D}^\infty$.

The key step of the proof of Theorem 3 is to prove the following Theorem 4, which is a stronger version of Theorem 2.

THEOREM 4. *Let* $\mathcal{D}_{\text{disc}}$ *be the class of discrete distributions that have positive probabilities only on a finite number of values. Let h be a distortion function defined on* $[0, 1]$ *and let* $\rho(\cdot)$ *be defined as in* (2) *with* $s = 1$. *Let* $\alpha_0 \in (0, 1)$ *be fixed. Then,* $\rho(\cdot)$ *has convex level sets with respect to* $\mathcal{D}_{\text{disc}} \cap \{F \mid q^-_{\alpha_0}(F) = q^+_{\alpha_0}(F)\}$ *if and only if one of the following four cases holds*:

(i) *There exists* $c \in [0, 1]$, *such that* $\rho = c\,\text{VaR}_0 + (1 - c)\,\text{VaR}_1$, *where* $\text{VaR}_0(F) := \inf\{x \mid F(x) > 0\}$ *and* $\text{VaR}_1(F) := \inf\{x \mid F(x) = 1\}$ (*i.e.,* $h(u) = 1 - c$, $\forall u \in (0, 1)$).

(ii) *There exists* $\alpha \in (0, 1)$ *such that* $\rho(F) = \text{VaR}_\alpha(F)$, $\forall F$ (*i.e.,* $h(u) = 1_{\{u > 1-\alpha\}}$).

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

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

where $q^-_\alpha(F) := \inf\{x \mid F(x) \geqslant \alpha\}$ *and* $q^+_\alpha(F) := \inf\{x \mid F(x) > \alpha\}$ (*i.e.,* $h(u) = (1 - c) \cdot 1_{\{u = 1-\alpha\}} + 1_{\{u > 1-\alpha\}}$).

(iv) $\rho(F) = \int x\, dF(x)$, $\forall F$ (*i.e.,* $h(u) = u$).

Furthermore, the risk measures ρ *listed in cases* (i)–(iv) *have convex level sets with respect to* $\mathcal{P}^\rho \supseteq \mathcal{D}^\infty$ ($\mathcal{P}^\rho$ *is defined in Theorem 3 and* $\mathcal{D}^\infty := \{F_X \mid X \in \mathcal{L}^\infty(\Omega, \mathcal{F}, P)\}$).

PROOF. See E-Companion EC.3. □

2.4. Co-Elicitability

The co-elicitability of $k \geqslant 2$ statistical functionals is a *weaker* notion of elicitability than the notion of elicitability of one statistical functional defined in Definition 1. The notion of co-elicitability is formulated in Lambert et al. (2008, Definition 9) as follows:

DEFINITION 3. $k \geqslant 2$ single-valued statistical functionals $\rho_1(\cdot), \ldots, \rho_k(\cdot)$ are called co-elicitable with respect to a class of distributions $\mathcal{P}$ if there exists a forecasting objective function $S\colon \mathbb{R}^{k+1} \to \mathbb{R}$ such that

$$(\rho_1(F), \ldots, \rho_k(F)) = \operatorname*{arg\,min}_{(x_1, \ldots, x_k)} \int S(x_1, \ldots, x_k, y)\, dF(y), \quad \forall F \in \mathcal{P}. \tag{13}$$

The notion of co-elicitability is weaker than that of elicitability because: (i) if for each $i = 1, \ldots, k$, ρ_i is elicitable with a corresponding forecasting objective function $S_i(\cdot, \cdot)$, then $(\rho_1, \ldots, \rho_k)$ are co-elicitable with the corresponding function S being defined as $S(x_1, \ldots, x_k, y) := \sum_{i=1}^k S_i(x_i, y)$; (ii) if $(\rho_1, \ldots, \rho_k)$ are co-elicitable, it does not imply that each ρ_i is elicitable.

Acerbi and Székely (2014) show that $(\text{VaR}_\alpha, \text{ES}_\alpha)$ are co-elicitable with respect to a class of distributions $\mathcal{P}$, which satisfy some restrictive conditions based on an intuitive argument; Fissler and Ziegel (2016) show that $(\text{VaR}_\alpha, \text{ES}_\alpha)$ are co-elicitable with respect to $\mathcal{P} = \{F \mid F$ has finite first moment and has unique α quantile$\}$, and the corresponding forecasting objective function S in Definition 3 may be specified as

$$\begin{aligned} S(x_1, x_2, y) = {}& (1_{\{x_1 > y\}} - \alpha)(-G_1(-x_1) + G_1(-y)) \\ & + \frac{1}{1-\alpha} G_2(-x_2) 1_{\{x_1 \leqslant y\}}(y - x_1) \\ & + G_2(-x_2)(x_1 - x_2) - \mathcal{G}_2(-x_2), \end{aligned} \tag{14}$$

where G_1 and G_2 are strictly increasing continuously differentiable functions, G_1 is F-integrable for any $F \in \mathcal{P}$, $\lim_{x \to -\infty} G_2(x) = 0$, and $\mathcal{G}_2' = G_2$, e.g., $G_1(x) = x$ and $G_2(x) = e^x$.

The co-elicitability of $(\text{VaR}_\alpha, \text{ES}_\alpha)$ implies that one can evaluate the performance of different forecasting procedures that forecast the *collection* of $(\text{VaR}_\alpha, \text{ES}_\alpha)$ by comparing their realized forecasting errors. More precisely, procedure 1 is considered to better forecast the *collection* of $(\text{VaR}_\alpha, \text{ES}_\alpha)$ than procedure 2 if

$$\frac{1}{T} \sum_{t=1}^T S(\text{var}_t^1, \text{es}_t^1, Y_t) < \frac{1}{T} \sum_{t=1}^T S(\text{var}_t^2, \text{es}_t^2, Y_t), \tag{15}$$

where $(\text{var}_t^i, \text{es}_t^i)$ are the forecasts generated by the ith procedure at time t, $i = 1, 2$, and Y_t is the realized loss at time t, $t = 1, \ldots, T$.

The co-elicitability of $(\text{ES}_\alpha, \text{VaR}_\alpha)$ does not lead to a reliable method for evaluating forecasts for ES_α. More precisely, even if procedure 1 better forecasts the collection

$(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ than procedure 2 *in the sense of* (15), procedure 1 may provide much worse forecast of ES_α than procedure 2; this is illustrated in Example 5 and Example 6 at the end of Section 2.5.2.

Theorem 1 (Theorem 3) identifies all general elicitable (elicitable) risk measures within the class of risk measures that satisfy Axioms A1–A5; a counterpart of the problem studied in Theorems 1 and 3 is the following one: For $k \geqslant 2$, can we identify all the k-tuple of risk measures $(\rho_1, \ldots, \rho_k)$ such that $(\rho_1, \ldots, \rho_k)$ are co-elicitable and each ρ_i satisfies Axioms A1–A5? Because co-elicitability is weaker than elicitability, the above problem is different from that studied in Theorems 1 and 3; the answer to the problem does not imply Theorem 1 or Theorem 3, and Theorem 1 or Theorem 3 does not provide a complete answer to the problem.

Some examples of risk measures that satisfy the conditions in the above open problem are provided in Fissler and Ziegel (2016). In addition to $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$, $(\mathrm{VaR}_{\alpha_1}, \ldots, \mathrm{VaR}_{\alpha_k}, \sum_{i=1}^k w_i \mathrm{ES}_{\alpha_i})$ are shown to be co-elicitable, where $0 < \alpha_1 < \cdots < \alpha_k < 1$, $(w_1, \ldots, w_k)$ are any weights satisfying $\sum_{i=1}^k w_i = 1$ and $w_i > 0$, $i = 1, \ldots, k$. However, the complete answer to the open problem is not known yet; we leave it for future research.

2.5. Backtesting a Risk Measure

As will be shown in the following subsections, there are three approaches for backtesting a risk measure: (i) the direct backtest, which tests if the point estimate or point forecast of the risk measurement under a model is equal to the unknown true risk measurement; (ii) the indirect backtest, which can be classified into two kinds: (a) the first kind of indirect backtests examine if the entire loss distribution, the entire tail loss distribution, or a collection of statistics including the risk measure of interest under a model are equal to the corresponding quantities under the true underlying unknown model; (b) the second kind of indirect backtests are based on the co-elicitability of a collection of risk measures; (iii) the forecast evaluation approach based on the elicitability of the risk measure.

We will also show in the subsections that (i) VaR and median shortfall can be backtested by all three approaches. (ii) There have been no direct back-testing methods for expected shortfall. (iii) Indirect backtesting methods for expected shortfall have been proposed in the literature. The first kind of indirect back testing for expected shortfall is a partial backtesting in the sense that (a) if an indirect backtesting for expected shortfall is not rejected, it will imply that the point forecast for expected shortfall will not be rejected; (b) however, if an indirect backtesting for expected shortfall is rejected, it will be unclear whether the point forecast for expected shortfall should be rejected. The second kind of indirect backtests that are based on the co-elicitability of $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ cannot answer the question whether the ES_α forecasted under a bank's model is more accurate than that forecasted under a benchmark model.

2.5.1. The Direct Backtesting Approach. The direct back-testing approach is to test whether the risk measurement calculated under a model is equal to the unknown true value of risk measurement. It concerns whether the point estimate or point forecast of the risk measure is acceptable or not. For example, suppose a bank reports that the $\mathrm{VaR}_{99\%}$ of its trading book is one billion. The direct backtesting approach answers the question whether the single number one billion is acceptable or not.

More precisely, suppose the loss of a bank on the tth day is L_t, $t = 1, 2, \ldots, T$. On each day $t-1$, the bank forecasts the risk measurement ρ of L_t based on the information available on day $t-1$, which is denoted as $\mathcal{F}_{t-1}$. Let $G_{t\mid t-1}$ denote the bank's model of the conditional distribution of L_t given $\mathcal{F}_{t-1}$, and let $\rho^{G_{t\mid t-1}}(L_t)$ denote the risk measurement of L_t under the model $G_{t\mid t-1}$. Suppose the unknown true conditional distribution of L_t given $\mathcal{F}_{t-1}$ is $F_{t\mid t-1}$ and the true risk measurement is denoted as $\rho^{F_{t\mid t-1}}(L_t)$. Then, the direct backtesting of the risk measure ρ is to test

$$\begin{aligned} &H_0\colon \rho^{G_{t\mid t-1}}(L_t) = \rho^{F_{t\mid t-1}}(L_t), \quad \forall\, t = 1, \ldots, T;\\ &H_1\colon \text{otherwise.} \end{aligned} \tag{16}$$

For $\rho = \mathrm{VaR}_\alpha$, the null hypothesis in (16) is equivalent to that $I_t := 1_{\{L_t > \mathrm{VaR}_\alpha(L_t)\}}$, $t = 1, \ldots, T$, are i.i.d. Bernoulli $(1-\alpha)$ random variables (Christoffersen 1998, Lemma 1). Based on such observation, Kupiec (1995) propose the proportion of failure test for backtesting VaR, which is closely related to the "traffic light" approach of back testing VaR adopted in the Basel Accord (Basel Committee on Banking Supervision 1996, 2006). Christoffersen (1998) proposes conditional coverage and independence tests for VaR within a firsft-order Markov process model. For more recent developments on the backtesting of VaR, see Lopez (1999a, 1998), Engle and Manganelli (2004), Christoffersen and Pelletier (2004), Haas (2005), Campbell (2006), Christoffersen (2010), Berkowitz et al. (2011), Gaglianone et al. (2011), Holzmann and Eulert (2014), etc.

As $\mathrm{MS}_\alpha = \mathrm{VaR}_{(1+\alpha)/2}$, the backtesting of median shortfall is exactly the same as that of VaR. In contrast, *there have been no direct back testing methods for expected shortfall* in the existing literature. The reason might be simple: The null hypothesis for direct back testing expected shortfall is that $\mathrm{ES}_\alpha^{G_{t\mid t-1}}(L_t) = \mathrm{ES}_\alpha^{F_{t\mid t-1}}(L_t)$. It might be *difficult* (*if not impossible*) to find a statistic whose distribution is known under the null hypothesis. In contrast, the distribution of the indicator random variable $I_t = 1_{\{L_t > \mathrm{VaR}_\alpha(L_t)\}}$ is known under the null hypothesis for direct back testing VaR, and hence I_t can be used to construct test statistic for direct back testing VaR.

2.5.2. The Indirect Backtesting Approach. There are two kinds of indirect backtesting approaches. The first kind of indirect backtesting approach concerns whether the bank's model of the *entire loss distribution* is the same as the unknown true loss distribution. More precisely, the indirect back-testing approach is to test

$$H_0\colon G_{t|t-1}(x) = F_{t|t-1}(x), \quad \forall x \in \mathbb{R}, \forall t = 1, \ldots, T; \qquad H_1\colon \text{otherwise}. \tag{17}$$

If the null hypothesis is not rejected, then it will imply that $\rho^{G_{t|t-1}}(L_t) = \rho^{F_{t|t-1}}(L_t)$, i.e., the risk measurement will not be rejected; however, if the null hypothesis is rejected, then it will be unclear whether the point forecast $\rho^{G_{t|t-1}}(L_t)$ should be rejected or not. Therefore, the kind of indirect back-testing approach can only serve as a *partial* back testing of a particular risk measure. For example, suppose a bank reports that the $\text{ES}_{99\%}$ of its trading book is one billion. Using the indirect backtesting approach, one can test the bank's model of the entire loss distribution. If the test is not rejected, then it will imply that the number one billion is acceptable; however, if the test is rejected, then it will be unclear if the number one billion should be accepted or rejected.

Strictly speaking, this indirect back-testing approach shall not be regarded as an approach for backtesting a particular risk measure, because the backtesting has nothing to do with any particular risk measure, although the test has partial implications on the acceptability of the point forecast of a particular risk measure.

This kind of indirect backtesting approach has been proposed for backtesting expected shortfall in the literature. Berkowitz (2001) proposes likelihood ratio tests based on censored Gaussian likelihood for the test (17). Kerkhof and Melenberg (2004) propose a functional delta method for testing the Hypothesis (17). Acerbi and Székely (2014) propose three indirect tests for back testing ES_α. The first two tests are to test the *entire tail loss distribution* under the assumption that VaR_α has already been tested and that $L_1, \ldots, L_T$ are independent:

$$H_0\colon G_{t|t-1,\alpha}(x) = F_{t|t-1,\alpha}(x), \quad \forall x \in \mathbb{R}, \forall t = 1, \ldots, T; \qquad H_1\colon \text{otherwise}, \tag{18}$$

where $G_{t|t-1,\alpha}$ and $F_{t|t-1,\alpha}$ and the α-tail distribution of $G_{t|t-1}$ and $F_{t|t-1}$, respectively, (see Example 2 for definition of α-tail distribution). The third test is the same as the test (17). All the three tests proposed by the authors require that one knows how to simulate random samples with distribution $G_{t|t-1}(\cdot)$ in order to simulate the test statistic and to calculate the p value of the test. Costanzino and Curran (2015) propose an approach to indirectly back test ES_α by testing

$$\begin{aligned} H_0\colon & \textstyle\int_\alpha^1 1_{\{L_t \leqslant \text{VaR}_p(L_t)\}}\,dp, \quad t = 1, \ldots, T, \text{ are i.i.d.}, \\ & \text{VaR}_p^{F_{t|t-1}}(L_t) = \text{VaR}_p^{G_{t|t-1}}(L_t), \quad \forall p \in [\alpha, 1), t = 1, \ldots, T \\ H_1\colon & \text{otherwise}. \end{aligned} \tag{19}$$

This approach does not need to simulate random samples under the null hypothesis in order to calculate the p value. McNeil and Frey (2000) assume that the loss process $\{L_t, t = 1, \ldots, T\}$ follows the dynamics $L_t = m_t + s_t Z_t$, where m_t and s_t are, respectively, the conditional mean and conditional standard deviation, and Z_t is a strict white noise. Under this assumption, they propose to back test ES_α by testing the hypothesis

$$\begin{aligned} H_0\colon & m_t^{G_{t|t-1}} = m_t, \quad s_t^{G_{t|t-1}} = s_t, \\ & \text{VaR}_\alpha^{G_{t|t-1}}(L_t) = \text{VaR}_\alpha^{F_{t|t-1}}(L_t), \\ & \text{ES}_\alpha^{G_{t|t-1}}(L_t) = \text{ES}_\alpha^{F_{t|t-1}}(L_t), \quad \forall t; \\ H_1\colon & \text{otherwise}. \end{aligned} \tag{20}$$

This test is an indirect test for ES_α because if the null hypothesis is rejected, it is not clear if the claim $\text{ES}_\alpha^{G_{t|t-1}}(L_t) = \text{ES}_\alpha^{F_{t|t-1}}(L_t), \forall t$ should be rejected or not.

The second kind of indirect back tests are those based on the co-elicitability of a collection of risk measures. For example, let $(\text{VaR}_\alpha^{\text{Ben}}(L_t), \text{ES}_\alpha^{\text{Ben}}(L_t)), t = 1, \ldots, T\}$, be the $(\text{VaR}_\alpha, \text{ES}_\alpha)$ forecasted under a benchmark model such as a standard model specified by the regulator. Fissler et al. (2015) propose the following two indirect backtests for backtesting ES_α:

$$\begin{aligned} H_0^-\colon & E_{t-1}\big[S(\text{VaR}_\alpha^{G_{t|t-1}}(L_t), \text{ES}_\alpha^{G_{t|t-1}}(L_t), L_t)\big] \\ & \geqslant E_{t-1}\big[S(\text{VaR}_\alpha^{\text{Ben}}(L_t), \text{ES}_\alpha^{\text{Ben}}(L_t), L_t)\big], \quad \forall t \\ H_1^-\colon & \text{otherwise}; \\ H_0^+\colon & E_{t-1}\big[S(\text{VaR}_\alpha^{G_{t|t-1}}(L_t), \text{ES}_\alpha^{G_{t|t-1}}(L_t), L_t)\big] \\ & \leqslant E_{t-1}\big[S(\text{VaR}_\alpha^{\text{Ben}}(L_t), \text{ES}_\alpha^{\text{Ben}}(L_t), L_t)\big], \quad \forall t \\ H_1^+\colon & \text{otherwise}, \end{aligned} \tag{21}$$

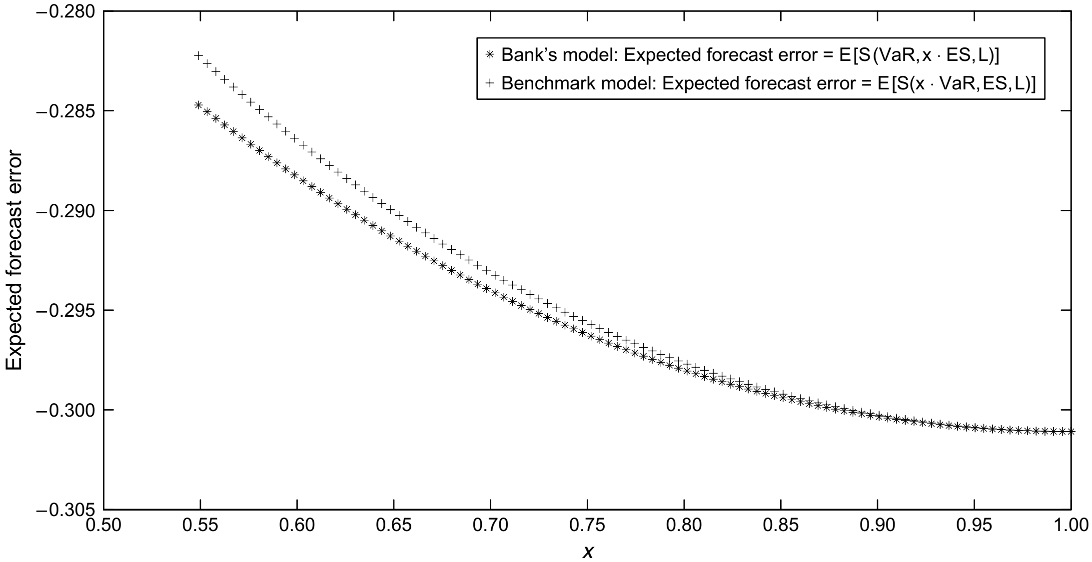
where $S(\cdot,\cdot,\cdot)$ is the forecasting objective function defined in (14) with $G_1(x) = x$ and $G_2(x) = e^x/(1+e^x)$.

These tests are indirect back tests for ES_α because no matter if these tests are rejected or not, we do now know whether $\text{ES}_\alpha^{G_{t|t-1}}$ is more accurate than $\text{ES}_\alpha^{\text{Ben}}(L_t)$. In fact, these tests are not able to find out which model gives a more accurate forecast for ES_α, as is shown in Example 5.

Example 5. Suppose the true distribution of a bank's loss random variable L is $N(\mu, \sigma^2)$. Let $\Phi_{\mu,\sigma}$ denote the distribution function of L. It can be easily calculated that

$$\begin{aligned} & E[S(x_1, x_2, L)] \\ = \; & x_1 \Phi_{\mu,\sigma}(x_1) - \left(\mu \Phi_{\mu,\sigma}(x_1) - \frac{\sigma}{\sqrt{2\pi}} e^{-(x_1-\mu)^2/(2\sigma^2)}\right) \\ & - \alpha(x_1 - \mu) + \frac{1}{1-\alpha} \frac{e^{-x_2}}{1+e^{-x_2}} \bigg[\mu - \bigg(\mu \Phi_{\mu,\sigma}(x_1) \\ & - \frac{\sigma}{\sqrt{2\pi}} e^{-(x_1-\mu)^2/(2\sigma^2)}\bigg) - x_1(1 - \Phi_{\mu,\sigma}(x_1))\bigg] \\ & + \frac{e^{-x_2}}{1+e^{-x_2}}(x_1 - x_2) - \log(1 + e^{-x_2}). \end{aligned}$$

Figure 1. A graph for the counterexample in Example 5.



Notes. The forecasting error of the bank's model (i.e., $\mathrm{E}[S(\mathrm{VaR}_\alpha, x\cdot \mathrm{ES}_\alpha, L)]$) is always smaller than that of the benchmark model (i.e., $\mathrm{E}[S(x\cdot \mathrm{VaR}_\alpha, \mathrm{ES}_\alpha, L)]$) for any $x \in (0.55, 1.0)$; therefore, the tests in (21) will conclude that the bank's model better forecasts ES_α than the benchmark model. However, the bank's model always underforecasts ES_α, while the benchmark model always truthfully forecasts ES_α. Such inconsistency, mainly because of the fact that co-elicitability does not imply elicitability, shows that the tests in (21) are not able to find out which model gives a more accurate forecast for ES_α. The lower boundary of x is 0.55 because the forecast $(\mathrm{VaR}_\alpha, x\cdot\mathrm{ES}_\alpha)$ given by the bank's model has to satisfy the condition $\mathrm{VaR}_\alpha < x\cdot \mathrm{ES}_\alpha$, which leads to $x > \mathrm{VaR}_\alpha / \mathrm{ES}_\alpha = 0.460/0.838 = 0.55$.

Let $\mu = -1.5$, $\sigma = 1.0$, and $\alpha = 0.975$, which is suggested in Basel Committee on Banking Supervision (2013). Then the true value of $(\mathrm{VaR}_\alpha(L), \mathrm{ES}_\alpha(L))$ is $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha) = (0.460, 0.838)$. Suppose the forecasts given by a bank's model are $(\mathrm{VaR}_\alpha, x\cdot\mathrm{ES}_\alpha)$ and those given by a benchmark model (preferred by the regulator) are $(x\cdot\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$, where $0 < x < 1$; hence, the bank's model always underforecasts ES_α but the benchmark model always truthfully forecasts ES_α; therefore, the bank's model should be rejected. However, these tests will conclude that the bank's model is better than the benchmark model because the forecasting error of the bank's model (i.e., $\mathrm{E}[S(\mathrm{VaR}_\alpha, x\cdot\mathrm{ES}_\alpha, L)]$) is always smaller than that of the benchmark model (i.e., $\mathrm{E}[S(x\cdot\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha, L)]$) for any $x \in (0.55, 1.0)$. In other words, even if the bank's model *underforecasts* the ES_α by as much as 45%, it will still be *wrongly* considered to be better than the benchmark model that truthfully forecasts ES_α, mainly because of the fact that co-elicitability does not imply elicitability, and some rather strange behavior of the forecasting objective function S defined in (14). This is illustrated by Figure 1.

Another drawback of these backtests is that the performance of the back tests further deteriorates when the scale of the loss random variable increases, because the term $G_2(-x_2)$ in Equation (14) goes to zero as x_2 goes to infinity. The consequence is that larger banks can more easily underreport ES than smaller banks if such backtests are used for back testing ES_α. This is illustrated in Example 6.

Example 6. Suppose there is a larger bank whose loss random variable is 15 times of the loss L in Example 5. Thus, the loss random variable of this larger bank has a normal distribution $N(\mu, \sigma^2)$ with $\mu = -1.5\times 15$, $\sigma = 15.0$. Let $\alpha = 0.975$. Note the true value of $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ is $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha) = (0.460, 0.838)\times 15$. Suppose the forecasts given by a bank's model are $(\mathrm{VaR}_\alpha, x\cdot\mathrm{ES}_\alpha)$ and those given by a benchmark model (preferred by the regulator) are $(x\cdot\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$. Again, as in Figure 1, Figure 2 shows that the back tests make the wrong conclusion on which model better forecasts ES_α. In addition, Figure 2 shows that the forecasting error for the bank's model almost remains unchanged when $x \in (0.55, 1.0)$, which is due to the fact that when ES_α is large enough, the term

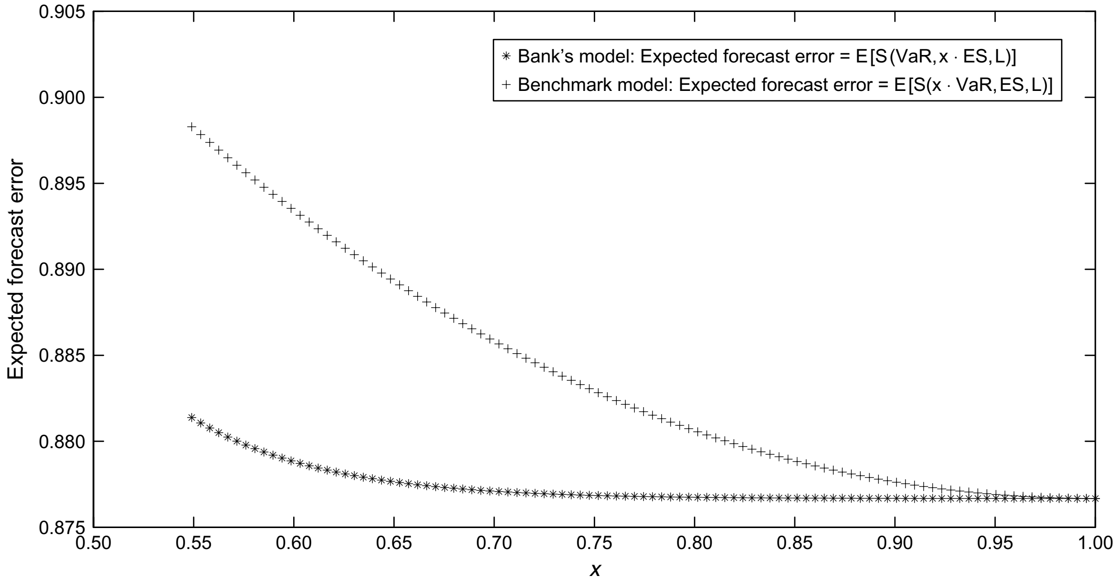
$$\mathrm{E}\big[(1/(1-\alpha))G_2(-x\cdot\mathrm{ES}_\alpha)1_{\{\mathrm{VaR}_\alpha<L\}}(L-\mathrm{VaR}_\alpha) + G_2(-x\cdot\mathrm{ES}_\alpha)(\mathrm{VaR}_\alpha - x\,\mathrm{ES}_\alpha) - \mathcal{G}_2(-x\cdot\mathrm{ES}_\alpha)\big]$$

in the expected forecasting error will be so small that the expected forecasting error will not change much when x varies. In other words, when the scale of the loss random variable L is large enough, the expected forecasting error $\mathrm{E}[S(\mathrm{VaR}_\alpha, x\cdot\mathrm{ES}_\alpha, L)]$ becomes insensitive to the value of x. This counterexample happens again *mainly because of some strange behavior of the forecasting objective function S defined in* (14).

2.5.3. The Backtesting Approach Based on the Elicitability of a Risk Measure. The backtesting approach based on the forecast evaluation framework and elicitability has been proposed to back test VaR. This approach requires a benchmark model because the elicitability concerns the comparison of multiple models rather than the validation of a single model. Lopez (1999a) proposes to define

Operations Research 64(5), pp. 1056–1072, © 2016 INFORMS

Figure 2. A graph for the counterexample in Example 6.



Notes. The expected forecasting error of the bank's model (i.e., $\mathrm{E}[S(\mathrm{VaR}_\alpha, x\cdot \mathrm{ES}_\alpha, L)]$) in Example 6 almost remains unchanged when $x\in(0.55, 1.0)$, because when ES_α is large enough, the term $\mathrm{E}[(1/(1-\alpha))G_2(-x\cdot\mathrm{ES}_\alpha)1_{\{\mathrm{VaR}_\alpha<L\}}(L-\mathrm{VaR}_\alpha)+G_2(-x\cdot\mathrm{ES}_\alpha)(\mathrm{VaR}_\alpha-x\,\mathrm{ES}_\alpha)-\mathscr{G}_2(-x\cdot\mathrm{ES}_\alpha)]$ in the expected forecasting error will be so small that the expected forecasting error will not change much when x varies. In other words, when the scale of the loss random variable L is large enough, the expected forecasting error $\mathrm{E}[S(\mathrm{VaR}_\alpha, x\cdot\mathrm{ES}_\alpha, L)]$ becomes insensitive to the value of x.

the forecasting error for VaR_α under the model $G_{t\,|\,t-1}$ as $\sum_{t=1}^{T} S(\mathrm{VaR}_\alpha^{G_{t|t-1}}(L_t), L_t)$, where $S(\cdot,\cdot)$ is a forecast objective function (loss function). Since VaR_α is elicitable with respect to $\mathscr{D}^1\cap\{F\mid q_\alpha^-(F)=q_\alpha^+(F)\}$, S can be defined as $S_\alpha(x,y)=(1_{\{x\geqslant y\}}-\alpha)(x-y)$. Then, the forecasting error is compared with a benchmark forecasting error calculated under a benchmark model to back test VaR_α.

In contrast, expected shortfall cannot be back tested by this approach because it is not elicitable, and therefore, no function S can be used to define the forecasting error.

3. Extension to Incorporate Multiple Models

The previous section addresses the issue of model uncertainty from the perspective of general 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, $i=1,\ldots,m$ on the state space $(\Omega,\mathscr{F})$. 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 satisfies Axioms A1–A5. Then by Lemma 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, $i=1,\ldots,m$. We then propose the following risk measure to incorporate multiple scenarios:

$$\rho(X)=f(\rho_1(X),\rho_2(X),\ldots,\rho_m(X)), \tag{22}$$

where $f\colon \mathbb{R}^m\to\mathbb{R}$ 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*: $f(a\tilde{x}+b\mathbf{1})=af(\tilde{x})+sb$, $\forall\,\tilde{x}\in\mathbb{R}^m$, $\forall\,a\geqslant 0$, $\forall\,b\in\mathbb{R}$, *where* $s>0$ *is a constant and* $\mathbf{1}:=(1,1,\ldots,1)\in\mathbb{R}^m$.

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

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

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, p. 145) to "capture the phenomenon of hedging;" it is used as one of the axioms for the maxmin expected utility that incorporates robustness.

Lemma 3. *A scenario aggregation function* $f\colon\mathbb{R}^m\to\mathbb{R}$ *satisfies Axioms B1–B3 if and only if there exists a set of* *weights* $\mathscr{W} = \{\tilde{w}\} \subset \mathbb{R}^m$ *with each* $\tilde{w} = (w_1, \ldots, w_m) \in \mathscr{W}$ *satisfying* $w_i \geqslant 0$ *and* $\sum_{i=1}^m w_i = 1$, *such that*

$$f(\tilde{x}) = s \cdot \sup_{\tilde{w}\in\mathscr{W}} \left\{ \sum_{i=1}^m w_i x_i \right\}, \quad \forall\, \tilde{x} \in \mathbb{R}^m. \tag{23}$$

Proof. First, we show that Axioms B1–B3 are equivalent to Kou et al. (2013, Axioms C1–C4) with $n_i = 1$, $i = 1, \ldots, m$. Axioms B1 and B2 are the same as the Axioms C1 and C2, respectively. Axiom C4 holds for any function when $n_i = 1$, $i = 1, \ldots, m$. Axioms C1 and C3 apparently imply 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

$$\begin{aligned} & f(\tilde{x}+\tilde{y}) - f(\tilde{x}) - f(\tilde{y}) \\ =\ & f\big(\tilde{x} - f(\tilde{x})/s + \tilde{y} - f(\tilde{y})/s\big) \\ =\ & 2f\big(\tfrac{1}{2}(\tilde{x} - f(\tilde{x})/s) + \tfrac{1}{2}(\tilde{y} - f(\tilde{y})/s)\big) \\ \leqslant\ & 2f(\tilde{x} - f(\tilde{x})/s) = 0. \end{aligned}$$

Hence, Axiom C3 holds. Therefore, Axioms B1–B3 are equivalent to Axioms C1–C4, and hence the conclusion of the lemma follows from (Kou et al. 2013, [Theorem 3.1]). □

In the representation (23), each weight $\tilde{w} \in \mathscr{W}$ 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.

Lemmas 1 and 3 lead to the following class of risk measures:[7]

$$\rho(X) = s \cdot \sup_{\tilde{w}\in\mathscr{W}} \left\{ \sum_{i=1}^m w_i \int X\, d(h_i \circ P_i) \right\}. \tag{24}$$

By Theorem 1, the requirement of general elicitability under each scenario leads to the following tail risk measure:

$$\rho(X) = s \cdot \sup_{\tilde{w}\in\mathscr{W}} \left\{ \sum_{i=1}^m w_i \,\mathrm{MS}_{i,\alpha_i}(X) \right\}, \tag{25}$$

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 (25) addresses the issue of model uncertainty and incorporates robustness from two aspects: (i) under each scenario i, MS_{i,α_i} is general elicitable with respect to $\mathscr{D}^1$ and statistically robust (Kou et al. 2006, 2013, Cont et al. 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 uses 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 $\rho_t(X_t, X_{t-1}, \ldots, X_{t-59}) := s_t \max\{(1/s_t)\,\mathrm{VaR}_{t-1}(X_t), (1/60)\sum_{i=1}^{60} \mathrm{VaR}_{t-i}(X_{t-i+1})\}$, where X_{t-i} is the trading book loss on the $(t-i)$th day; $s_t \geqslant 3$ is a constant that is specified by the regulator based on the back-testing result of the institution's VaR model; $\mathrm{VaR}_{t-i}(X_{t-i+1})$ 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 61st model under which $X = 0$ with probability one. Assume that the trading book composition and the size of the positions remain the same over the 60-day periods. Then, $X_t, X_{t-1}, \ldots, X_{t-59}$ can be regarded as the realization of the same random loss under different distributions. In such cases, the Basel II risk measure is a special case of the class of risk measures considered in (25); it incorporates 61 models and two priors: one is $\tilde{w} = (1/s, , \ldots, , 1-1/s)$, the other $\tilde{w} = (1/60, /60, \ldots, 1/60, 0)$. 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 (25).

In a consultative document released by the Bank for International Settlement (Basel Committee on Banking Supervision 2013, p. 3), 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 $\rho_t(X_t, X_{t-1}, \ldots, X_{t-59}) := s\max\{(1/s)\,\mathrm{ES}_{t-1}(X_t),\ (1/60)\sum_{i=1}^{60}\mathrm{ES}_{t-i}(X_{t-i+1})\}$, where $\mathrm{ES}_{t-i}(X_{t-i+1})$ is the ES at 97.5% confidence level calculated on day $t-i$, $i = 1, \ldots, 60$. Assume that the trading book composition and the size of the positions remain the same over the 60-day periods. Then, the proposed Basel 3.5 risk measure is a special case of the class of risk measures considered in (24).[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 not 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 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 the 99% level, or, alternatively, MS at the 99% level.

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 November 26, 2012.

	ES			MS			
α (%)	$ES_{\alpha,1}$	$ES_{\alpha,2}$	$ES_{\alpha,2}-ES_{\alpha,1}$	$MS_{\alpha,1}$	$MS_{\alpha,2}$	$MS_{\alpha,2}-MS_{\alpha,1}$	$\frac{ES_{\alpha,2}-ES_{\alpha,1}}{MS_{\alpha,2}-MS_{\alpha,1}}-1$
97.0	19,956	21,699	1,743	19,070	19,868	798	118.4%
97.5	20,586	22,690	2,104	19,715	20,826	1,111	89.3%
98.0	21,337	23,918	2,581	20,483	22,011	1,529	68.8%
98.5	22,275	25,530	3,254	21,441	23,564	2,123	53.3%
99.0	23,546	27,863	4,317	22,738	25,807	3,070	40.6%
99.5	25,595	32,049	6,454	24,827	29,823	4,996	29.2%

Notes. $ES_{\alpha,i}$ and $MS_{\alpha,i}$ 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., $ES_{\alpha,2}-ES_{\alpha,1}$) is much larger than that of MS (i.e., $MS_{\alpha,2}-MS_{\alpha,1}$).

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

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

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

- 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 the 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 $ES_{\alpha,i}$ and $MS_{\alpha,i}$ 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., $ES_{\alpha,2}-ES_{\alpha,1}$) is much larger than that of MS (i.e., $MS_{\alpha,2}-MS_{\alpha,1}$), indicating that ES is more sensitive to model misspecification than MS.

5. Comments

5.1. Criticism of Value-at-Risk

As pointed out by Aumann and Serrano (2008, p. 813), "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: (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 et al. (2013, Section 6.1) for more discussion. (3) Although subadditivity ensures that $\rho(X_1)+\rho(X_2)$ is an upper bound for $\rho(X_1+X_2)$, this upper bound may not be valid in the face of model uncertainty.[10] (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 much larger than the latter.[11] (5) Subadditivity is not necessarily needed for capital allocation or asset allocation.[12] (6) It is often argued that if a nonsubadditive 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 entire 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.

Even if one believes in subadditivity, VaR (and median shortfall) satisfies subadditivity in most relevant situations. In fact, Daníelsson et al. (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 et al. 2011, p. 150). Garcia et al. (2007, p. 483) 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."

(iv) Embrechts et al. (2015, p. 763) argue that "ES are more robust than VaR according to the new notion of robustness concerning the sensitivity of a risk measure to the uncertainty of dependence in risk aggregation." because VaR is not aggregation robust but expected shortfall is. However, their counterexample (i.e., their Example 2.2) 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;[13] note that by Corollary 3.7 of Cont et al. (2010) expected shortfall is not Hampel robust.

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

Theorems 1 and 3 generalize the main result in Ziegel (2014), 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 (Theorem 3.1) for risk measures with convex acceptance set $\mathcal{N}$ and convex rejection set $\mathcal{N}^c$ under 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 ψ: $\mathbb{R} \to [1, \infty)$. 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 general elicitability and hence the topological conditions may not hold. For example, the results in Bellini and Bignozzi (2015), which rely on the characterization theorem in Weber (2006), make strong assumptions on the forecasting objective function $S(\cdot, \cdot)$,[14] requiring a more restrictive definition of elicitability than Gneiting (2011). The elicitability of a risk measure is also related to the statistical theory for the evaluation of probability forecasts (Lai et al. 2011).

The axioms in this paper are based on economic considerations. Other axioms based on mathematical considerations include convexity (Föllmer and Schied 2002, Frittelli and Gianin 2002, 2005), comonotonic subadditivity (Song and Yan 2006, 2009; Kou et al. 2006, 2013), and comonotonic convexity (Song and Yan 2006, 2009).

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/opre.2016.1539.

Acknowledgments

The authors thank the editor, Rama Cont, and two anonymous referees for their insightful comments that help to improve the paper. The research of Steven Kou is supported in part by the Singapore Ministry of Education [Grant MOE2014-T2-1-006] and the Natural Science Foundation of China [Grant R-2015-S-005]. Xianhua Peng is supported in part by the University Grants Committee of HKSAR of China and School of Science of Hong Kong University of Science and Technology [Projects DAG12SC05-3, SBI14SC08, and R9307].

Endnotes

1. The axioms used in Wang et al. (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, $x_n = \rho(x_n) \geqslant \rho(\lambda) \geqslant \rho(y_n) = y_n$. Letting $n \to \infty$ yields $\rho(\lambda) = \lambda$, $\forall\, \lambda \in \mathbb{R}$; hence, Axiom A3 holds. (ii) By the monotonic axiom, $\rho(\min(X, M)) \leqslant \rho(\min(\max(X, -M), M)) \leqslant$

$\rho(\max(X,-M))$. 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 $\lambda_n\rho(X^M)-\epsilon=\rho(\lambda_n X^M-\epsilon)\leqslant\rho(\lambda X^M)\leqslant\rho(\lambda_n X^M+\epsilon)=\lambda_n\rho(X^M)+\epsilon$. Letting $n\to\infty$ yields $\lambda\rho(X^M)-\epsilon\leqslant\rho(\lambda X^M)\leqslant\lambda\rho(X^M)+\epsilon,\forall\epsilon>0$. Letting $\epsilon\downarrow 0$ leads to $\rho(\lambda X^M)=\lambda\rho(X^M)$, $\forall\lambda\geqslant 0$. Letting $M\to\infty$ and applying Axiom A5 result in $\rho(\lambda X)=\lambda\rho(X)$, $\forall\lambda\geqslant 0$. 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 $P(X>x)\geqslant P(Y>x)$ for all x, which implies that for a risk measure ρ represented by (2), $\rho(X)\geqslant\rho(Y)$.

3. The term "median shortfall" is also used in Moscadelli (2004) and So and Wong (2012) but is, respectively, defined as median$[X\mid X>u]$ for a constant u and median$[X\mid X>\mathrm{VaR}_\alpha(X)]$, which are different from that defined in Kou et al. (2013). In fact, the definition in the aforementioned second paper is the same as the "tail conditional median" proposed in Kou et al. (2006).

4. Indeed, for $\alpha\in(0,1)$, by definition,

$$\begin{aligned}\mathrm{MS}_\alpha(X)&=\inf\{x\mid F_{\alpha,X}(x)\geqslant\tfrac{1}{2}\}\\&=\inf\{x\mid (F_X(x)-\alpha)/(1-\alpha)\geqslant\tfrac{1}{2}\}\\&=\inf\{x\mid F_X(x)\geqslant(1+\alpha)/2\}=\mathrm{VaR}_{(1+\alpha)/2}(X);\end{aligned}$$

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

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 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 $h(x):=1_{\{x\geqslant 1-\alpha\}}$, but it can be shown that q_α^+ cannot be represented by (5). 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 $X_1=X_0\cdot 1_{\{X_0\leqslant F_{X_0}^{-1}(\alpha)\}}+(X_0+c)\cdot 1_{\{X_0>F_{X_0}^{-1}(\alpha)\}}$. 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 $\rho_\Delta(X_0)=q_\alpha^+(X_0)$.

6. We thank an anonymous referee for pointing this out to us.

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

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 et al. (2013). The natural risk statistics are axiomatized by a different set of axioms including a comonotonic subadditivity axiom.

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. In fact, suppose we are concerned with obtaining an upper bound for $\mathrm{ES}_\alpha(X_1+X_2)$. In practice, because of 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 $\widehat{\mathrm{ES}}_\alpha(X_1)+\widehat{\mathrm{ES}}_\alpha(X_2)<\mathrm{ES}_\alpha(X_1)+\mathrm{ES}_\alpha(X_2)$.

11. 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 strike \$95. Then the margin requirement for X_1+X_2, $\rho(X_1+X_2)$, should not be larger than \$5, as $X_1+X_2\leqslant 5$. 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 $\rho(X_1+X_2)$.

12. Kou et al. (2013, Section 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), Peng et al. (2013), Xi et al. (2014), and the references therein for asset allocation methods using VaR and Basel Accord risk measures.

13. Aggregation robustness is a notion of robustness that is weaker than Hampel robustness. By Huber and Ronchetti (2009, Theorem 2.21), a risk measure (statistical functional) ρ is Hampel-robust at a distribution F is essentially equivalent to that ρ is weakly continuous at F. More precisely, if ρ is Hampel robust at F, then for any $\epsilon>0$, there exists $\delta>0$ such that for $\forall G\in\mathcal{N}_\delta(F):=\{H\mid d(F,H)<\delta\}$, it holds that $|\rho(F)-\rho(G)|<\epsilon$. In contrast, ρ is aggregation robust at F means that for any $\epsilon>0$, there exists $\delta>0$ such that for $\forall G\in\mathcal{N}_\delta(F)\cap\mathcal{A}_F$, it holds that $|\rho(F)-\rho(G)|<\epsilon$, where $\mathcal{A}_F:=\{H\mid$ there exist integer $m>0$ and random variables $X_1,\ldots,X_m,X_1',\ldots,X_m'$, such that $X_i\overset{d}{\sim}X_i'$, $i=1,\ldots,m$, $\sum_{i=1}^m X_i\overset{d}{\sim}F$, and $\sum_{i=1}^m X_i'\overset{d}{\sim}H.\}$. Since $\mathcal{N}_\delta(F)\cap\mathcal{A}_F\subsetneq\mathcal{N}_\delta(F)$, aggregation robustness is weaker than Hampel robustness.

14. 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 $\epsilon>0$, $S(x,y)\leqslant\psi(y)$ for some gauge function ψ.

References

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

Acerbi C, Székely B (2014) Back-testing expected shortfall. *Risk Magazine* (November 17), 76–81.

Adrian T, Shin HS (2014) Procyclical leverage and Value-at-Risk. *Rev. Financial Stud.* 27(2):373–403.

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

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

Basel Committee on Banking Supervision (1996) Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements. 1-15. Document, Bank for International Settlements, Basel, Switzerland.

Basel Committee on Banking Supervision (2006) International convergence of capital measurement and capital standards: A revised framework. Document, Bank for International Settlements, Basel, Switzerland, p. 48.

Basel Committee on Banking Supervision (2009) Revisions to the Basel II market risk framework. Document, Bank for International Settlements, Basel, Switzerland, p. 15.

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, p. 18.

Bellini F, Bignozzi V (2015) On elicitable risk measures. *Quant. Finance* 15(5):725–733.

Berkowitz J (2001) Testing density forecasts, applications to risk management. *J. Bus. Econom. Statist.* 19(4):465–474.

Berkowitz J, Christoffersen P, Pelletier D (2011) Evaluating value-at-risk models with desk-level data. *Management Sci.* 57(12):2213–2227.

Brunnermeier MK, Crockett A, Goodhart C, Persaud AD, Shin HS (2009) *The Fundamental Principles of Financial Regulation: 11th Geneva Report on the World Economy* (Centre for Economic Policy Research, Washington, DC).

Brunnermeier MK, Pedersen LH (2009) Market liquidity and funding liquidity. *Rev. Financial Stud.* 22(6):2201–2238.

Campbell SD (2006) A review of backtesting and backtesting procedures. *J. Risk* 9(2):1–17.

Cherny A, Madan D (2009) New measures for performance evaluation. *Rev. Financial Stud.* 22(7):2571–2606.

Christoffersen P (1998) Evaluating interval forecasts. *Internat. Econom. Rev.* 39(4):841–862.

Christoffersen P (2010) Backtesting. Cont RB, ed. *Encyclopedia of Quantitative Finance* (John Wiley & Sons, Hoboken, NJ).

Christoffersen PF, Pelletier D (2004) Backtesting value-at-risk: A duration-based approach. *J. Financial Econometrics* 2(1):84–108.

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

Costanzino N, Curran M (2015) Backtesting general spectral risk measures with application to expected shortfall. Preprint, University of Toronto and Bank of Montreal.

Daníelsson J, Jorgensen BN, Samorodnitsky G, Sarma M, de Vries CG (2013) Fat tails, VaR and subadditivity. *J. Econometrics* 172(2): 283–291.

Davis MHA (2013) Consistency of risk measure estimates. Preprint, Imperial College London.

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

Duffie D, Pan J (1997) An overview of value at risk. *J. Derivatives* 4(3): 7–49.

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

Embrechts P, Wang B, Wang R (2015) Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance Stochastics* 19(4):763–790.

Engelberg J, Manski CF, Williams J (2009) Comparing the point predictions and subjective probability distributions of professional forecasters. *J. Bus. Econom. Statist.* 27(1):30–41.

Engle RF, Manganelli S (2004) CAViaR: Conditional autoregressive value at risk by regression quantiles. *J. Bus. Econom. Statist.* 22(4): 367–381.

Fama EF, Miller MH (1972) *The Theory of Finance* (Dryden Press, Hinsdale, IL).

Fissler T, Ziegel JF (2016) Higher order elicitability and Osband's principle. *Ann. Statist.* 44(4):1680–1707.

Fissler T, Ziegel JF, Gneiting T (2015) Expected shortfall is jointly elicitable with value at risk—Implications for backtesting. *Risk Magazine* (December).

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

Foster DP, Hart S (2009) An operational measure of riskiness. *J. Political Econom.* 117(5):785–814.

Foster DP, Hart S (2013) A wealth-requirement axiomatization of riskiness. *Theoret. Econom.* 8(2):591–620.

Frittelli M, Gianin ER (2002) Putting order in risk measures. *J. Banking Finance* 26(7):1473–1486.

Frittelli M, Gianin ER (2005) Law invariant convex risk measures. *Adv. Math. Econom.* 7:33–46.

Gaglianone WP, Lima LR, Linton O, Smith DR (2011) Evaluating value-at-risk models via quantile regression. *J. Bus. Econom. Statist.* 29(1): 150–160.

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

Gilboa I, Schmeidler D (1989) Maxmin expected utility with non-unique prior. *J. Math. Econom.* 18(2):141–153.

Glasserman P, Kang W (2013) Design of risk weights. *Oper. Res.* 62(6): 1204–1220.

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

Gneiting T (2011) Making and evaluating point forecasts. *J. Amer. Statist. Assoc.* 106(494):746–762.

Gordy MB (2003) A risk-factor model foundation for ratings-based bank capital rules. *J. Financial Intermediation* 12(3):199–232.

Gordy MB, Howells B (2006) Procyclicality in Basel II: Can we treat the disease without killing the patient? *J. Financial Intermediation* 15(3):395–417.

Haas M (2005) Improved duration-based backtesting of value-at-risk. *J. Risk* 8:17–38.

Hansen LP (2013) Challenges in identifying and measuring systemic risk. Brunnermeier M, Krishnamurthy A, eds. *Risk Topography: Systemic Risk and Macro Modeling* (University of Chicago Press, Chicago), 15–30.

Hansen LP, Sargent TJ (2001) Robust control and model uncertainty. *Amer. Econom. Rev.* 91(2):60–66.

Hansen LP, Sargent TJ (2007) *Robustness* (Princeton University Press, Princeton, NJ).

Hart S (2011) Comparing risks by acceptance and rejection. *J. Political Econom.* 119(4):617–638.

Holzmann H, Eulert M (2014) The role of the information set for forecasting–with applications to risk management. *Ann. Appl. Statist.* 8(1):595–621.

Huber PJ (1981) *Robust Statistics* (John Wiley & Sons, New York).

Huber PJ, Ronchetti EM (2009) *Robust Statistics*, 2nd ed. (John Wiley & Sons, Hoboken, NJ).

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

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

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

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

Kerkhof J, Melenberg B (2004) Backtesting for risk-based regulatory capital. *J. Banking Finance* 28:1845–1865.

Kou S, Peng X (2014) On the measurement of economic tail risk. Preprint, version February 2014, Arxiv: http://arxiv.org/abs/1401.4787v2.

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

Kou S, Peng X, Heyde CC (2013) External risk measures and Basel Accords. *Math. Oper. Res.* 38(3):393–417.

Kupiec P (1995) Techniques for verifying the accuracy of risk management models. *J. Derivatives* 3(2):73–84.

Lai TL, Shen D, Gross S (2011) Evaluating probability forecasts. *Ann. Statist.* 39(5):2356–2382.

Lambert NS, Pennock DM, Shoham Y (2008) Eliciting properties of probability distributions. *Proc. 9th ACM Conf. Electronic Commerce, EC '08* (ACM, New York), 129–138.

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

Lopez JA (1999a) Methods for evaluating value-at-risk estimates. *Federal Reserve Bank of San Francisco Econom. Rev.* 2(2):3–17.

Lopez JA (1998) Regulatory evaluation of value-at-risk models. *J. Risk* 1(2):37–64.

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

McNeil A, Frey R (2000) Estimation of tail-related risk measures for heteroskedastic financial time series: An extreme value approach. *J. Empirical Finance* 7:271–300.

McNeil AJ, Frey R, 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 KH (1985) *Providing Incentives for Better Cost Forecasting*. PhD thesis, University of California at Berkeley.

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

Rockafellar RT, Uryasev S (2002) Conditional value-at-risk for general loss distributions. *J. Banking Finance* 26(7):1443–1471.

Savage LJ (1971) Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* 66(336):783–810.

Schmeidler D (1986) Integral representation without additivity. *Proc. Amer. Math. Soc.* 97(2):255–261.

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

Shi Z, Werker BJM (2012) Short-horizon regulation for long-term investors. *J. Banking Finance* 36(12):3227–3238.

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

Song Y, Yan J-A (2006) The representation of two types of functionals on $\mathscr{L}^\infty(\Omega, \mathscr{F})$ and $\mathscr{L}^\infty(\Omega, \mathscr{F}, \mathbb{P})$. *Sci. China Ser. A: Math.* 49(10): 1376–1382.

Song Y, Yan J-A (2009) Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. *Insurance: Math. Econom.* 45(3):459–465.

Tasche D (2002) Expected shortfall and beyond. *J. Banking Finance* 26(7): 1519–1533.

Thomson W (1979) Eliciting production possibilities from a well-informed manager. *J. Econom. Theory* 20(3):360–380.

Von Hippel PT (2005) Mean, median, and skew: Correcting a textbook rule. *J. Statist. Ed.* 13(2).

Wang R, Ziegel JF (2015) Elicitable distortion risk measures: A concise proof. *Statist. Probab. Lett.* 100(May):172–175.

Wang SS, Young VR, Panjer HH (1997) Axiomatic characterization of insurance prices. *Insurance: Math. Econom.* 21(2):173–183.

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

Winkler RL, Jose VRR (2011) Scoring rules. Cochran J, ed. *Wiley Encyclopedia of Operations Research and Management Science*, Vol 7 (John Wiley & Sons, New York), 4733–4744.

Xi J, Coleman TF, Li Y, Tayal A (2014) A gradual non-convexification method for minimizing VaR. *J. Risk* 16(3):23–47.

Xia J (2013) Comonotonic convex preferences. Preprint, Academy of Mathematics and Systems Science, Chinese Academy of Sciences.

Ziegel JF (2014) Coherence and elicitability. *Math. Finance*, ePub ahead of print September 3, http://dx.doi.org/10.1111/mafi.12080.

Steven Kou is the director of the Risk Management Institute and a Provost's Chair Professor of Mathematics at the National-University of Singapore. His research interests are in quantitative finance, applied probability, and statistics. He is the winner of the 2002 Erlang Prize awarded by the Applied Probability Society of INFORMS. He is currently the Vice Chair of Applied Probability for the Financial Service Section of INFORMS.

Xianhua Peng is an assistant professor at the Department of Mathematics of Hong Kong University of Science and Technology. His research focuses on risk management, financial engineering, and applied probability, in particular risk measures, credit derivatives, real estate derivatives, and market microstructure.