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

To cite this article:

Full terms and conditions of use: http://pubsonline.informs.org/page/terms-and-conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article’s accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org
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 elicitation (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 elicitation and co-elicitation; in particular, we show that the co-elicitation 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; elicitation; 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 expected 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 elicitation of a risk measure is a property based on a decision-theoretic framework for evaluating the performance of different forecasting procedures (Gneiting 2011). The elicitation 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.
Kou and Peng: On the Measurement of Economic Tail Risk
Operations Research 64(5), pp. 1056–1072, © 2016 INFORMS

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 on which 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 on which 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 on which 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 on which 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 on which 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 on which 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 on which 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.
A risk measure \( \varphi : \mathcal{E} \to \mathbb{R} \) satisfies Axioms A1–A5 if and only if there exists a distortion function \( h(\cdot) \) such that

\[
\rho(X) = s \int X d(h \circ P) = \frac{s}{\mathbb{E}(h(X))} \int_{-\infty}^{\infty} h(X) dF(sX), \quad \forall X \in \mathcal{E},
\]

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

**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 \( \lim_{s \to 0} h((X - d)^+) = h(X^+) \) 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 \( \alpha \) is defined as the left \( \alpha \)-quantile of \( F_\alpha \):

\[
\text{VaR}_\alpha(X) := q_\alpha(F) := F^{-1}_\alpha(\alpha) = \inf\{x \mid F(x) \geq \alpha\}. \tag{3}
\]

For \( \alpha = 0 \), VaR of \( X \) at level \( \alpha \) is defined to be \( \text{VaR}_0(X) := \inf\{x \mid F(x) > 0\} \) and \( \text{VaR}_0(X) \) is equal to the essential infimum of \( X \). For \( \alpha \in (0, 1] \), \( \rho \) in (2) with \( s = 1 \) is equal to \( \text{VaR}_\alpha \) if \( h(x) := 1_{\{x > 0\}}; \rho \) in (2) with \( s = 1 \) is equal to \( \text{VaR}_\alpha \) if \( h(x) := 1_{\{\alpha \leq 0\}} \). 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 \( \alpha \) is defined as the mean of the \( \alpha \)-tail distribution of \( X \) (Tasche 2002, Rockafellar and Uryasev 2002), i.e.,

\[
\text{ES}_\alpha(X) := \frac{1}{\alpha} \int_{-\infty}^{\infty} x dF_{\alpha,X}(x), \quad \alpha \in (0, 1],
\]

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

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

**Lemma 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 \( \mathcal{E} \supseteq \mathcal{E}^\infty(\Omega, \mathcal{F}, P) \) be a set of random variables \((\mathcal{E} \text{ may include unbounded random variables})\).
For $\alpha = 1$, ES of $X$ at level $\alpha$ is defined as $\text{ES}_\alpha(X) := F^{-1}_X(1)$. If the distribution $F_X$ is continuous, then $F_{a,X}$ is the same as the conditional distribution of $X$ given that $X > \text{VaR}_a(X)$; if $F_X$ is not continuous, then $F_{a,X}(x)$ is a slight modification of the conditional loss distribution. For $\alpha \in [0, 1)$, $\rho(X)$ in (2) with $s = 1$ is equal to $\text{ES}_\alpha(X)$ if

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

For $\alpha = 1$, $\rho(X)$ in (2) with $s = 1$ is equal to $\text{ES}_\alpha(X)$ if $h(x) = 1_{[1-0]}$.

**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)

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

For $\alpha = 1$, MS at level $\alpha$ is defined as $\text{MS}_\alpha(X) := F^{-1}_X(1)$.

Therefore, MS at level $\alpha$ can capture the tail risk and considers both the size and likelihood of losses beyond the VaR at level $\alpha$, because it measures the median of the loss size conditional on that the loss exceeds the VaR at level $\alpha$. It can be shown that

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

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

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

$$F_{a,n}(x) := \begin{cases} 0, & x < c \\ (1 - \beta(n))(1 - e^{-\lambda(x-c)}) & \quad \text{for } x \leq c \\ + \beta(n)1_{[c,x]} & \quad \text{for } x > c, \end{cases}$$

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

In other words, $F_{a,n}$ is the mixture of $c + \exp(\lambda)$ (with probability $(1 - \beta(n))$) and the point mass $\delta_n$ (with probability $\beta(n)$). Under $F_{a,n}$, a large loss with size $n$ occurs with a small probability $\beta(n)$. For each $n$, $\text{ES}_{a,n}$, which is the mean of $F_{a,n}$, is always equal to $c + \mu + 1/\lambda$; hence, $\text{ES}_\alpha$ fails in the same way as $\text{MS}_\alpha$ regarding the detection of the large loss with size $n$, which may occur beyond $\text{VaR}_{1/(1-\alpha)}$. This example shows that the degree to which $\text{ES}_\alpha$ quantifies the risk beyond $\text{VaR}_{1/(1-\alpha)}$ might also be limited. After all, $\text{MS}_\alpha$ and $\text{ES}_\alpha$ are, respectively, the median and the mean of the same $\alpha$-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

$$\rho_\alpha(X) := \int_{(0,1]} F^{-1}_X(u) d\Delta(u),$$

where $\Delta$ 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. A special case of (5) is the spectral risk measure (Acerbi 2002, Definition 3.1), defined as

$$\rho(X) = \int_{(0,1]} F^{-1}_X(u)\phi(u) du,$$

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

Because of the requirement that $\phi$ 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 $\phi$ in (6) is an increasing function. The MINMVAR 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 $\alpha \geq 0$ 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 $\rho$ satisfies Axiom A4 (law invariance), then $\rho(X)$ only depends on $F_X$; hence, $\rho$ 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 $\rho$. 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 ρ(X) (or ρ(Fx)) summarizes the distribution Fx by a real number ρ(X), just as a point forecast for X does. In practice, the true distribution Fx is unknown and one has to estimate the unknown true value ρ(Fx). As one may come up with different procedures to forecast ρ(Fx), it is an important issue to evaluate which procedure provides a better forecast of ρ(Fx).

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 Fy. The expected forecasting error is given by

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

where $S(x, y): \mathbb{R}^2 \rightarrow \mathbb{R}$ is a forecasting objective function, e.g., $S(x, y) = (x - y)^2$ or $S(x, y) = |x - y|$. The optimal point forecast corresponding to $S$ is

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

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

A statistical functional $\rho$ 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 $\rho$ is elicitable, then one can evaluate two point forecasting methods by comparing their respective expected forecasting error $ES(x, Y)$. As $F_y$ is unknown, the expected forecasting error can be approximated by the average $(1/n) \sum_{i=1}^n S(x_i, Y_i)$, 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 $\rho$ 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 $S: \mathbb{R}^2 \rightarrow \mathbb{R}$ such that

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

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

$$\mathcal{D}^\alpha := \{F \mid F$ is a distribution on $\mathbb{R}$ and has finite first moment$, \quad (8)$$

and the corresponding forecasting objective function can be defined as

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

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 $\alpha$ is defined as $VaR_\alpha := q_\alpha^-(\cdot)$, i.e., the left $\alpha$-quantile. As a single-valued statistical functional, $VaR_\alpha = q_\alpha^-(\cdot)$ is elicitable with respect to $\mathcal{D}^\alpha \cap \{F \mid q_\alpha^-(F) = q_\alpha^+(F)\}$ but not elicitable with respect to $\mathcal{D}^\alpha$, 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 $F$, $VaR_\alpha(F) = arg\min_x \int S_\alpha(x, y) dF(y)$ holds for all $\alpha \in (0, 1)$ except for a countable set of $\alpha$ 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 $S: \mathbb{R}^2 \rightarrow \mathbb{R}$ such that

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

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 $\text{VaR}_\alpha$ is general elicitable with respect to $\mathcal{D}$, which is consistent with the fact that $\alpha$-quantile as a set-valued statistical functional is elicitable (in the sense of Definition 1) with respect to $\mathcal{D}$ 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 $\mathcal{D}_{\text{disc}}$ 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 $\mathcal{D}_{\text{disc}}$; 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 $\alpha$, provides a precise description of the average size of loss beyond $\text{VaR}_\alpha$ 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 $L \geq 0.6$.

**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 $\rho : \mathcal{D} \rightarrow \mathbb{R}$ be a risk measure that satisfies Axioms A1–A5 and $\mathcal{D} \supseteq \mathcal{D}_{\text{disc}}(\Omega, \mathcal{F}, P)$. Let $\mathcal{P}_\rho := \{F_X | X \in \mathcal{D}\}$ and let $\mathcal{D}_{\text{disc}}$ 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 $\mathcal{D}_{\text{disc}}$ if and only if one of the following two cases holds:

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

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

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

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 $\rho$ to be general elicitable with respect to $\mathcal{D}$ is that $\rho$ has convex level sets with respect to $\mathcal{D}$, i.e., $\rho(F)$ is convex for all $F \in \mathcal{D}$. The second and the key step is to prove the following theorem.

**Theorem 2.** 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 $s = 1$. Then, $\rho(\cdot)$ has convex level sets with respect to $\mathcal{D}_{\text{disc}}$ if and only if one of the following four cases holds:

(i) There exists $c \in [0, 1]$ such that $\rho = c \text{VaR}_\alpha + (1-c) \text{VaR}_\beta$, where $\text{VaR}_\alpha(F) := \inf\{x | F(x) > 0\}$ and $\text{VaR}_\beta(F) := \inf\{x | F(x) = 1\}$ (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., $h(u) = 1_{[\alpha, 1]}$).

(iii) There exists $\alpha \in (0, 1)$ and $c \in [0, 1)$ such that $\rho(F) = c q_\alpha(F) + (1-c) q_\alpha(F)$, $\forall F$,

$$\text{where } q_\alpha(F) := \inf\{x | F(x) \geq \alpha\} \text{ and } q_\alpha(F) := \inf\{x | F(x) > \alpha\} (i.e., h(u) = (1 - c) \cdot 1_{[\alpha, 1]} + 1_{(0, \alpha]}).$$

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

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

**Proof.** See E-Companion EC.2. □

**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 general-
ization 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 \(\rho: \mathbb{R} \to \mathbb{R}\) be a risk measure that satisfies Axioms A1–A5 and \(\mathcal{D} \subseteq \mathcal{D}^\infty (\Omega, \mathcal{F}, P)\). Let \(\mathcal{P}^\rho := \{F_X | X \in \mathcal{X}\}\) 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 \(\mathcal{D}_{\text{disc}} \cap \{F | q^+_{\alpha_0}(F) = q^+_{\alpha_0}(F)\}\) if and only if one of the following two cases holds:

1. \(\rho = \text{Var}_{\alpha_0} = q^+_{\alpha_0}\).
2. \(\rho(F) = \int x \, dF(x), \forall F\).

Furthermore, \(\text{Var}_{\alpha_0}\) and the mean functional are elicitable (in the sense of Definition 1) with respect to a larger class \(\mathcal{D}' \cap \{F | q^+_{\alpha_0}(F) = q^+_{\alpha_0}(F)\}\). 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 \(\mathcal{D}_{\text{disc}} \cap \{F | q^+_{\alpha_0}(F) = q^+_{\alpha_0}(F)\}\) if and only if one of the following four cases holds:

1. There exists \(c \in [0, 1]\), such that \(\rho = c\text{Var}_{\alpha_0} + (1 - c)\text{VaR}_{\alpha_0}\), where \(\text{Var}_{\alpha_0}(F) := \inf\{x | F(x) > 0\}\) and \(\text{VaR}_{\alpha_0}(F) := \inf\{x | F(x) = 1\}\) (i.e., \(h(u) = 1 - c\), \(\forall u \in (0, 1)\)).
2. There exists \(\alpha \in (0, 1)\) such that \(\rho(F) = \text{VaR}_{\alpha_0}(F), \forall F\) (i.e., \(h(u) = 1_{[\alpha, 1 - \alpha]}\)).
3. There exists \(\alpha \in (0, 1)\) and \(c \in [0, 1]\) such that \(\rho(F) = c q^+_{\alpha_0}(F) + (1 - c) q^+_{\alpha_0}(F), \forall F\),
   \[\text{(12)}\]
   where \(q^+_{\alpha_0}(F) := \inf\{x | F(x) > \alpha\}\) and \(q_{\alpha_0}^-(F) := \inf\{x | F(x) > \alpha\}\) (i.e., \(h(u) = 1 - c\) \(\cdot 1_{[\alpha, 1 - \alpha]} + 1_{[1 - \alpha, 1]}\)).
4. \(\rho(F) = \int x \, dF(x), \forall F\) (i.e., \(h(u) = u\)).

Furthermore, the risk measures \(\rho\) 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 \(\mathcal{D}^\infty := \{F_X | X \in \mathcal{X} (\Omega, \mathcal{F}, P)\}\).

**Proof.** See E-Companion EC.3. □

### 2.4. Co-Elicitability

The co-elicitability of \(k \geq 2\) statistical functionals is a **weaker** notion of elicitation 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.** \(k \geq 2\) 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 \(S: \mathbb{R}^{k+1} \to \mathbb{R}\) such that

\[
S(x_1, \ldots, x_k, y) = \min_{\{x_1, \ldots, x_k\}} \int \sum_{i=1}^k S(x_i, y) \, dF(y), \quad \forall F \in \mathcal{P}.
\]

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

Acerbi and Székeley (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 \(\mathcal{P} = \{F | F\) has finite first moment and has unique \(\alpha\) quantile\}, and the corresponding forecasting objective function \(S\) in Definition 3 may be specified as

\[
S(x_1, x_2, y) = (1_{[y, +\infty]} - \alpha) (-G_1(x_1) + G_1(-y)) + \frac{1}{1 - \alpha} G_2(-y_2) 1_{[y_2, +\infty]} (y_1 - x_1) + G_2(-y_2) (x_1 - x_2) - G_2(-x_2),
\]

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

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

where \((\text{var}_t^1, \text{es}_t^1)\) are the forecasts generated by the \(i\)th 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 \(\text{ES}_{\alpha}\). More precisely, even if procedure 1 better forecasts the collection...
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 \( k \geq 2 \), 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 \( \rho_i \) satisfies Axioms A1–A5? Because co-elicitation is weaker than elicitation, 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 \((\text{VaR}_a, \text{ES}_a)\), \((\text{VaR}_a, \ldots, \text{VaR}_a, \sum_{i=1}^k w_i \text{ES}_a)\) 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 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-elicitation of a collection of risk measures; (iii) the forecast evaluation approach based on the elicitation 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.

### 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 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 \( r \)th day is \( L_r \), \( t = 1, 2, \ldots, T \). On each day \( t = 1 \), the bank forecasts the risk measurement \( \rho \) of \( L_r \) based on the information available on day \( t = 1 \), which is denoted as \( \mathcal{F}_{t=r-1} \). Let \( G_{r=r-1} \) denote the bank’s model of the conditional distribution of \( L_r \) given \( \mathcal{F}_{t=r-1} \), and let \( \rho^{G_{r=r-1}}(L_r) \) denote the risk measurement of \( L_r \) under the model \( G_{r=r-1} \). Suppose the unknown true conditional distribution of \( L_r \) given \( \mathcal{F}_{t=r-1} \) is \( F_{r=r-1} \) and the true risk measurement is denoted as \( \rho^{F_{r=r-1}}(L_r) \). Then, the direct backtesting of the risk measure \( \rho \) is to test

\[
H_0: \rho^{G_{r=r-1}}(L_r) = \rho^{F_{r=r-1}}(L_r), \quad \forall t = 1, \ldots, T;
H_1: \text{otherwise.}
\]

For \( \rho = \text{VaR}_a \), the null hypothesis in (16) is equivalent to that \( I_t := 1_{\{L_t > \text{VaR}_a(L_t)\}}, t = 1, \ldots, T \), are i.i.d. Bernoulli \((1 - \alpha)\) random variables (Christoffersen 1998, Lemma 1). Based on such observation, Kupiec (1995) proposes 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 first-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 (2000), Christoffersen (2010), Berkowitz et al. (2011), Gaglianone et al. (2011), Holzmann and Eulert (2014), etc.

As \( MS_a = \text{VaR}_a(t+a/2) \), 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 \( \text{ES}_a^{G_{r=r-1}}(L_r) = \text{ES}_a^{F_{r=r-1}}(L_r) \). 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 \( I_t := 1_{\{L_t > \text{VaR}_a(L_t)\}} \) 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
\[ H_0: \quad G_{t|t-1}(x) = F_{t|t-1}(x), \quad \forall x \in \mathbb{R}, \forall t = 1, \ldots, T; \]
\[ H_1: \quad \text{otherwise.} \tag{17} \]

If the null hypothesis is not rejected, then it will imply that \( p_{G(t)}(L_t) = p_{F(t)}(L_t) \), 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 \( p_{G(t)}(L_t) \) 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 ES\(_\text{99\%}\) 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 backtesting ES\(_\alpha\). The first two tests are to test the entire tail loss distribution under the assumption that VaR\(_\alpha\) has already been tested and that \( L_1, \ldots, L_T \) are independent:
\[ H_0: \quad G_{t|t-1, \alpha}(x) = F_{t|t-1, \alpha}(x), \quad \forall x \in \mathbb{R}, \forall t = 1, \ldots, T; \]
\[ H_1: \quad \text{otherwise,} \tag{18} \]

where \( G_{t|t-1, \alpha} \) and \( F_{t|t-1, \alpha} \) and the \( \alpha\)-tail distribution of \( G_{t|t-1} \) and \( F_{t|t-1} \), respectively, (see Example 2 for definition of \( \alpha\)-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\(_\alpha\) by testing
\[ H_0: \quad \int_0^1 \mathbf{1}_{\{x \leq \text{VaR}(t)\}} dp, \quad t = 1, \ldots, T, \quad \text{are i.i.d.,} \]
\[ \text{VaR}_{p(t)}(L_t) = \text{VaR}_{p(t)}^{G_{t|t-1}}(L_t), \quad \forall p \in [0, 1], t = 1, \ldots, T \tag{19} \]
\[ H_1: \quad \text{otherwise.} \]

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\(_\alpha\) by testing the hypothesis
\[ H_0: \quad m_t^{G_{t|t-1}} = m_t, \quad s_t^{G_{t|t-1}} = s_t, \]
\[ \text{VaR}_{p(t)}^{G_{t|t-1}}(L_t) = \text{VaR}_{p(t)}^{F_{t|t-1}}(L_t), \quad \forall t; \tag{20} \]
\[ H_1: \quad \text{otherwise.} \]

This test is an indirect test for ES\(_\alpha\) because if the null hypothesis is rejected, it is not clear if the claim ES\(_{G_{t|t-1}}(L_t) = \text{ES}_{F_{t|t-1}}(L_t), \forall t \) 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 \( (\text{VaR}_{p}^{\text{Ben}}(L_t), \text{ES}_{p}^{\text{Ben}}(L_t)), t = 1, \ldots, T \), be the (VaR\(_\alpha\), 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\(_\alpha\):
\[ H_0: \quad E_{t-1}[S(\text{VaR}_{p}^{G_{t|t-1}}(L_t), \text{ES}_{p}^{G_{t|t-1}}(L_t), L_t)] \geq E_{t-1}[S(\text{VaR}_{p}^{\text{Ben}}(L_t), \text{ES}_{p}^{\text{Ben}}(L_t), L_t)], \quad \forall t; \tag{21} \]
\[ H_1: \quad \text{otherwise,} \]
\[ H_0^*: \quad E_{t-1}[S(\text{VaR}_{p}^{G_{t|t-1}}(L_t), \text{ES}_{p}^{G_{t|t-1}}(L_t), L_t)] \leq E_{t-1}[S(\text{VaR}_{p}^{\text{Ben}}(L_t), \text{ES}_{p}^{\text{Ben}}(L_t), L_t)], \quad \forall t; \]
\[ H_1^*: \quad \text{otherwise,} \]

where \( S(\cdot, \cdot, \cdot) \) is the forecasting objective function defined in (14) with \( G(\cdot) = x \) and \( G(\cdot) = x^4/(1 + e^x) \).

These tests are indirect back tests for ES\(_\alpha\) because no matter if these tests are rejected or not, we do now know whether ES\(_{G_{t|t-1}}\) is more accurate than ES\(_{\text{Ben}}\). In fact, these tests are not able to find out which model gives a more accurate forecast for ES\(_\alpha\), 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
\[ E[S(x_1, x_2, L)] = x_1 \Phi_{\mu, \sigma}(x_1) - \left( \mu \Phi_{\mu, \sigma}(x_1) - \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \right) \]
\[ - \alpha(x_1 - \mu) + \frac{1}{1 - \alpha} \left( 1 + e^{-x_2} \right) \left[ \mu - \left( \mu \Phi_{\mu, \sigma}(x_1) - \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \right) - x_1 (1 - \Phi_{\mu, \sigma}(x_1)) \right] \]
\[ + \frac{\sqrt{2\pi}}{1 - \alpha} e^{-x_2} (x_1 - x_2) - \log(1 + e^{-x_2}). \]
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 $(\text{VaR}_a(L), \text{ES}_a(L))$ is $(\text{VaR}_a, \text{ES}_a) = (0.460, 0.838)$. Suppose the forecasts given by a bank’s model are $(\text{VaR}_a, x \cdot \text{ES}_a)$ and those given by a benchmark model (preferred by the regulator) are $(x \cdot \text{VaR}_a, \text{ES}_a)$, where $0 < x < 1$; hence, the bank’s model always underestimates $\text{ES}_a$, but the benchmark model always truthfully forecasts $\text{ES}_a$; 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., $E[S(\text{VaR}_a, x \cdot \text{ES}_a, L)]$) is always smaller than that of the benchmark model (i.e., $E[S(x \cdot \text{VaR}_a, \text{ES}_a, L)]$) for any $x \in (0.55, 1.0)$; therefore, the tests in (21) conclude that the bank’s model better forecasts $\text{ES}_a$ than the benchmark model. However, the bank’s model always underestimates $\text{ES}_a$, while the benchmark model always truthfully forecasts $\text{ES}_a$. 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 $\text{ES}_a$.

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 $(\text{VaR}_a, \text{ES}_a)$ is $(\text{VaR}_a, \text{ES}_a) = (0.460, 0.838) \times 15$. Suppose the forecasts given by a bank’s model are $(\text{VaR}_a, x \cdot \text{ES}_a)$ and those given by a benchmark model (preferred by the regulator) are $(x \cdot \text{VaR}_a, \text{ES}_a)$. Again, as in Figure 1, Figure 2 shows that the back tests make the wrong conclusion on which model better forecasts $\text{ES}_a$. 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 $\text{ES}_a$ is large enough, the term

\[ E[(1/(1-\alpha))G_2(-x \cdot \text{ES}_a)1_{\text{VaR}_a < L}(L - \text{VaR}_a) + G_2(-x \cdot \text{ES}_a)(\text{VaR}_a - x \cdot \text{ES}_a) - G_2(-x \cdot \text{ES}_a)] \]

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 $E[S(\text{VaR}_a, x \cdot \text{ES}_a, L)]$ 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).

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 $\text{ES}_a$. 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.
the forecasting error for VaR \( \alpha \) under the model \( G_{ij}^{-1} \) as 
\[ \sum_{i=1}^{T} S(\text{VaR}_{\alpha} G_{ij}^{-1}(L_i), L_i), \]
where \( S(\cdot, \cdot) \) is a forecast objective function (loss function). Since VaR \( \alpha \) is elicitable with respect to \( \mathcal{F} \cap \{ F \mid q_{\alpha}^-(F) = q_{\alpha}^+(F) \} \), \( S \) can be defined as 
\[ S_{\alpha}(x, y) = (1 - \alpha)(x - y). \]
Then, the forecasting error is compared with a benchmark forecasting error calculated under a benchmark model to back test VaR \( \alpha \).

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 \( \Omega, \mathcal{F} \). Each \( P_i \) corresponds to one model or one scenario, which may refer to a specific economic regime such as an economic boom and a financial crisis. The loss distribution of a random loss \( X \) under different scenarios can be substantially different. For example, the VaR calculated under the scenario of the 2007 financial crisis is much higher than that under a scenario corresponding to a normal market condition due to the difference of loss distributions.

Suppose that under the \( i \)th scenario, the measurement of risk is given by \( \rho_i \), that satisfies Axioms A1–A5. Then by Lemma 1, \( \rho_i \) 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:

\[
\rho(X) = f(\rho_1(X), \rho_2(X), \ldots, \rho_m(X)),
\]

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

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

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

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

**Axiom B3. Uncertainty aversion:**
\[
f(\tilde{x}) = f(\tilde{y}), \quad \text{then for any } \alpha \in (0, 1), \ f(\alpha \tilde{x} + (1 - \alpha)\tilde{y}) \leq f(\tilde{x}).
\]

Axiom B1 states that if the risk measurement of \( Y \) is an affine function of that of \( X \) under each scenario, then the aggregate risk measurement of \( Y \) is also an affine function of that of \( X \). Axiom B2 states that if the risk measurement of \( X \) is less than or equal to that of \( Y \) under each scenario, then the aggregate risk measurement of \( X \) is also less than or equal to that of \( Y \). Axiom B3 is proposed by Gilboa and Schmeidler (1989, 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 \( f : \mathbb{R}^m \to \mathbb{R} \) satisfies Axioms B1–B3 if and only if there exists a set of
weights $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^m$ with each $\tilde{w} = (w_1, \ldots, w_m) \in \mathcal{W}$ satisfying $w_i \geq 0$ and $\sum_{i=1}^{m} w_i = 1$, such that

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

**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

$$
f(\tilde{x} + \tilde{y}) - f(\tilde{x}) - f(\tilde{y})
= f(\tilde{x} - f(\tilde{x})/s + \tilde{y} - f(\tilde{y})/s)
= 2f(\frac{1}{s}(\tilde{x} - f(\tilde{x})/s) + \frac{1}{s}(\tilde{y} - f(\tilde{y})/s))
\leq 2f(\tilde{x} - f(\tilde{y})/s) = 0.
$$

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

In the representation (23), each weight $\tilde{w} \in \mathcal{W}$ can be regarded as a prior probability on the set of scenarios; more precisely, $w_i$ can be viewed as the likelihood that the scenario $i$ happens.

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

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

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

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

where $\text{MS}_{i, \alpha_i}(X)$ is the median shortfall of $X$ at confidence level $\alpha_i$ calculated under the $i$th scenario (model). The risk measure $\rho$ in (25) addresses the issue of model uncertainty and incorporates robustness from two aspects: (i) under each scenario $i$, $\text{MS}_{i, \alpha_i}$ is general elicitable with respect to $\mathcal{G}^i$ and statistically robust (Kou et al. 2006, 2013, Cont et al. 2010); (ii) $\rho$ 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 $t$th day is specified as $p_t(X_t, X_{t-1}, \ldots, X_{t-59}) := s \cdot \max\{(1/s) \cdot \text{VaR}_{t-1}(X_t), (1/60) \cdot \sum_{i=1}^{60} \text{VaR}_{t-1}(X_{t+i})\}$, where $X_{t-i}$ is the trading book loss on the $(t-i)$th day; $s \geq 3$ is a constant that is specified by the regulator based on the back-testing result of the institution’s VaR model; $\text{VaR}_{t-1}(X_{t+i})$ is the 10-day VaR at 99% confidence level calculated on day $t-i$, which corresponds to the $i$th model, $i = 1, \ldots, 60$. Define the 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 $t$th day is defined as $p_t(X_t, X_{t-1}, \ldots, X_{t-59}) := s \cdot \max\{(1/s) \cdot \text{ES}_{t-1}(X_t), (1/60) \cdot \sum_{i=1}^{60} \text{ES}_{t-1}(X_{t+i})\}$, where $\text{ES}_{t-1}(X_{t+i})$ 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). \(\text{ES}\) 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.

<table>
<thead>
<tr>
<th>α (%)</th>
<th>ES(_{\alpha,1})</th>
<th>ES(_{\alpha,2})</th>
<th>ES(<em>{\alpha,2} - ES(</em>{\alpha,1})</th>
<th>MS(_{\alpha,1})</th>
<th>MS(_{\alpha,2})</th>
<th>MS(<em>{\alpha,2} - MS(</em>{\alpha,1})</th>
<th>ES(<em>{\alpha,2} - ES(</em>{\alpha,1})) / MS(<em>{\alpha,2} - MS(</em>{\alpha,1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>97.0</td>
<td>19,956</td>
<td>21,699</td>
<td>1,743</td>
<td>19,070</td>
<td>19,868</td>
<td>798</td>
<td>118.4%</td>
</tr>
<tr>
<td>97.5</td>
<td>20,586</td>
<td>22,690</td>
<td>2,104</td>
<td>19,715</td>
<td>20,826</td>
<td>1,111</td>
<td>39.3%</td>
</tr>
<tr>
<td>98.0</td>
<td>21,337</td>
<td>23,918</td>
<td>2,581</td>
<td>20,483</td>
<td>22,011</td>
<td>1,529</td>
<td>68.8%</td>
</tr>
<tr>
<td>98.5</td>
<td>22,275</td>
<td>25,530</td>
<td>3,254</td>
<td>21,441</td>
<td>23,564</td>
<td>2,123</td>
<td>53.3%</td>
</tr>
<tr>
<td>99.0</td>
<td>23,546</td>
<td>27,863</td>
<td>4,317</td>
<td>22,738</td>
<td>25,807</td>
<td>3,070</td>
<td>40.6%</td>
</tr>
<tr>
<td>99.5</td>
<td>25,595</td>
<td>32,049</td>
<td>6,454</td>
<td>24,827</td>
<td>29,823</td>
<td>4,996</td>
<td>29.2%</td>
</tr>
</tbody>
</table>

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

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

- Model 2. The same as model 1 except that the conditional distribution is specified as \( \epsilon_t \sim t_{\nu} \), where \( t_{\nu} \) denotes \( t \) distribution with degree of freedom \( \nu \).

We, respectively, fit the two models to the historical data of daily returns of the S&P 500 Index during 11/26/1980–11/26/2012 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\(_{\alpha}\) and MS\(_{\alpha}\) 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\(_{\alpha}\).

However, the median shortfall at level α does address this issue by measuring the median size of the tail loss distribution beyond VaR\(_{\alpha}\).

(ii) There is a pathological counterexample that, for some level α, the VaR\(_{\alpha}\) 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 α > 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, Danielsson 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 \( \alpha \) such that \( F^{-1}(\cdot) \) is not continuous at \( \alpha \). There are only at most a countable number of such \( \alpha \); in fact, if \( F \) is a continuous distribution, then no such \( \alpha \) exists. On the contrary, for any other \( \alpha \), VaR at level \( \alpha \) is aggregation robust, because VaR at level \( \alpha \) 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 \( ES_\alpha > VaR_\alpha \). This argument is misleading because ES at level \( \alpha \) should be compared with VaR at level \((1 + \alpha)/2\) (i.e., MS at level \( \alpha \)). ES_\alpha may be smaller (i.e., less conservative) than MS_\alpha, as mean may be smaller than median. For example, if the tail loss distribution is a Weibull distribution with a shape parameter larger than 3.44, then ES_\alpha is smaller than MS_\alpha (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 \( N \) and convex rejection set \( N^c \) under two topological conditions on \( N \): (1) there exists \( x \in \mathbb{R} \) with \( \delta_x \in N \) such that for \( y \in \mathbb{R} \) and \( \delta_y \in N^c \), \( (1 - \alpha)\delta_x + \alpha\delta_y \in N \) for sufficiently small \( \alpha > 0 \); (2) \( N \) is \( \psi \)-weakly closed for some gauge function \( \psi: \mathbb{R} \rightarrow [1, \infty) \). 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 \( S(\cdot, \cdot) \), 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](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 National 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 additive 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 \( \rho(X) = \lambda \rho(X) \). Since \( \rho(\lambda) = \rho(\lambda) \), it follows that \( \rho(\lambda) = \lambda \). (ii) The same reasoning implies \( \rho(\lambda Y) = \lambda \rho(Y) \). Hence, \( \rho(\lambda) = \lambda \). Letting \( n \rightarrow \infty \) yields \( \rho(\lambda) = \lambda \), \( \forall \lambda \in \mathbb{R} \); hence, Axiom A3 holds. (ii) By the monotonic axiom, \( \rho(\min(X, M)) \leq \rho(\min(\max(X, M), M)) \leq \rho(M) \).
\(\rho(\max(X, -M))\). Letting \(M \to \infty\) and using the conditions \(\rho(\min(X, M)) \to \rho(X)\) and \(\rho(\max(X, -M)) \to \rho(X)\) as \(M \to \infty\) in their continuity axiom, without need of the condition \(\lim_{d \to 0^+} \rho(x - d^+) = \rho(x^+)\). Axiom A5 follows. (iii) We then show positive homogeneity holds, i.e., \(\rho(\lambda X) = \lambda \rho(X)\) for any \(X\) and any \(\lambda > 0\). For any \(X\) and \(M > 0\), denote \(X^M := \min(\max(X, -M), M)\). For any \(\epsilon > 0\) and \(\lambda > 0\), there exist \(\{\lambda_n\} \subset \mathbb{Q}^+\) such that \(\lambda_n \to \lambda\) as \(n \to \infty\) and \(\lambda_n \rho(X^M) - \epsilon = \rho(\lambda_n X^M - \epsilon) \leq \rho(\lambda X^M) \leq \rho(\lambda_n X^M + \epsilon) = \lambda_n \rho(X^M) + \epsilon\). Letting \(n \to \infty\) yields \(\lambda \rho(X^M) - \epsilon \leq \rho(\lambda X^M) \leq \lambda \rho(X^M) + \epsilon, \forall \epsilon > 0\). Letting \(\epsilon \downarrow 0\) leads to \(\rho(\lambda X^M) = \lambda \rho(X^M), \forall \lambda > 0\). Letting \(M \to \infty\) and applying Axiom A5 result in \(\rho(\lambda X) = \lambda \rho(X), \forall \lambda > 0\). Their comonotonic additivity axiom and positive homogeneity imply Axiom A1.

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

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[\(X \mid X > \text{Var}_R(X)\)], 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,

\[
\text{MS}_\alpha(x) = \inf \{x \mid F_\alpha(x) = \alpha\} = \inf \{x \mid \frac{F_\alpha(x) - \alpha}{1 - \alpha} = \frac{1}{2}\} = \text{VaR}_{\frac{1}{2} \alpha}(x);
\]

for \(\alpha = 1\), by definition, \(\text{MS}_1(x) = F_{\frac{1}{2}}(1) = \text{VaR}_1(x)\); for \(\alpha = 0\), by definition, \(F_0(x) = F(x)\) and hence \(\text{MS}_0(x) = F_0(x) = \text{VaR}_0(x)\).

5. In fact, for any fixed \(u \in (0, 1)\), \(F_{\frac{1}{2}}(u) = \text{VaR}_u(x)\) 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, \(\text{VaR}_u(x)\) satisfies monotonicity, positive homogeneity, and comonotonic additivity, which implies that \(p_u\) satisfies Axioms A1–A4 for any \(\alpha\). On \(\mathcal{L}^\infty(\Omega, \mathcal{F}, P), p_u\) automatically satisfies Axiom A5. On the other hand, for an \(\alpha \in (0, 1)\), the right quantile \(q^u_\alpha(x):= \inf \{x \mid F(x) > \alpha\}\) is a special case of the risk measure defined in (2) with \(h(x)\) being defined as \(h(x) := 1_{[0, 1] \setminus [0, \alpha]}\), but it can be shown that \(q^u_\alpha(x)\) cannot be represented by (5).

Indeed, suppose for the sake of contradiction that there exists a \(\Delta\) such that \(q^u_\alpha(x) = \rho^u(X)\), \(\forall X \in \mathcal{L}^\infty(\Omega, \mathcal{F}, P)\). Let \(X_0\) have a strictly positive density on its support. Then, \(F_{\frac{1}{2}}(u)\) is continuous and strictly increases on \(\{0, 1\}\). Let \(\epsilon > 0\) be a constant. Define \(X_1 = X_0 \cdot 1_{\{X_0 \leq F_{\frac{1}{2}}(u)\}} + (X_0 + \epsilon) \cdot 1_{\{X_0 > F_{\frac{1}{2}}(u)\}}\). It follows that \(q^u_\alpha(X_1) > q^u_\alpha(X_0)\) that \(\alpha(1, 1) = 1\), which in combination with the strict monotonicity of \(F_{\frac{1}{2}}(u)\) implies that \(p_u(X_1) = p_u(X_0)\). Consequently, \(\rho^u(X_1) = \rho^u(X_0)\), implying \(\rho^u(X_0) = q^u_\alpha(X_0)\).

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

7. Gilboa and Schmeidler (1989) consider inf \(\int u(x) dP\) without \(\gamma\); 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 used the same set of axioms.

References


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


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.