ANNUAL REVIEWS

Annual Review of Statistics and Its Application

Risk Measures: Robustness, Elicitability, and Backtesting

Xue Dong He,[1] Steven Kou,[2] and Xianhua Peng[3]

[1]Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong, China; email: xdhe@se.cuhk.edu.hk

[2]Questrom School of Business, Boston University, Boston, MA 02215, USA; email: kou@bu.edu

[3]HSBC Business School, Peking University, Shenzhen 518055, China; email: xianhuapeng@phbs.pku.edu

Annu. Rev. Stat. Appl. 2022. 9:9.1–9.26

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-030718-105122

Keywords

risk measures, robustness, elicitability, backtesting, value-at-risk, expected shortfall, median shortfall

Abstract

Risk measures are used not only for financial institutions' internal risk management but also for external regulation (e.g., in the Basel Accord for calculating the regulatory capital requirements for financial institutions). Though fundamental in risk management, how to select a good risk measure is a controversial issue. We review the literature on risk measures, particularly on issues such as subadditivity, robustness, elicitability, and backtesting. We also aim to clarify some misconceptions and confusions in the literature. In particular, we argue that, despite lacking some mathematical convenience, the median shortfall—that is, the median of the tail loss distribution—is a better option than the expected shortfall for setting the Basel Accords capital requirements due to statistical and economic considerations such as capturing tail risk, robustness, elicitability, backtesting, and surplus invariance.

1. INTRODUCTION

The research question in this article is how to select good risk measures to appraise financial risk. Obviously, to evaluate the appropriateness of risk measures, one has to specify the objectives first. In terms of objectives, risk measures can be classified into two categories: internal risk measures used for internal risk management and asset allocation at individual institutions and external risk measures used for external regulation or margin requirements and imposed for relevant financial institutions and traders.

A risk measure may be suitable for internal management but not for external regulation, and vice versa. This article focuses on external risk measures and reviews three strands of related research: robustness, elicitability, and backtesting.

A major difficulty in measuring 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." A risk measure is said to be robust if it can accommodate model misspecification and it has statistical robustness with respect to changes in the data. Robustness is especially important for an external risk measure used for banking regulation, as the risk measure needs to be consistently implemented across all relevant financial institutions.

Different procedures may be used to forecast a 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. This is a desirable property because the embedded forecasting objective function can then be used for evaluating different forecasting procedures.

Elicitability is closely related to backtesting, whose objective is to evaluate the performance of a risk forecasting model. If a risk measure is elicitable, then the sample average forecasting error based on the objective function can be used for backtesting the risk measure. An external risk measure used for banking regulation needs to be backtested, as the regulator needs to evaluate the quality of the model and procedure used in forecasting the external risk measure.

The outline of the article is as follows. In Section 2, we review the axiomatic approaches of risk measures and acceptance sets, along with some important examples including value-at-risk (VaR), expected shortfall (ES), and median shortfall (MS). Section 3 discusses the controversial subadditivity and convexity axioms, which are the most important axioms for the coherent and convex risk measure, respectively, and have been at the center of the debate on what risk measure is a good one. Robustness, elicitability, and backtesting are studied in Sections 4–6, respectively. Section 7 presents the Basel Accord risk measures.

2. RISK MEASURES AND ACCEPTANCE SETS

We consider a risk manager who wants to propose a risk measure or an acceptance set for her particular use. The risk manager can be a regulator, a clearing house, or a financial institution. She represents uncertainty in the market by a measurable space $(\Omega, \mathcal{F})$. Due to model uncertainty, the risk manager may have a set of probability models, which are represented by $\mathcal{P}$, a class of probability measures on $(\Omega, \mathcal{F})$. We denote with $\mathcal{L}^0(\Omega, \mathcal{F}, \mathcal{P})$ the set of all random variables, which is a vector space. We say that $X, Y \in \mathcal{L}^0(\Omega, \mathcal{F}, \mathcal{P})$ are identically distributed if X and Y have the same distribution under any $\mathbb{P} \in \mathcal{P}$.

In many existing papers on risk measures, $\mathcal{P}$ is either the set of all probability measures on $(\Omega, \mathcal{F})$ (and thus the risk manager does not have a probabilistic model) or a singleton (and thus there is no model uncertainty). However, in practice, the risk measure of a capital position or a portfolio's P&L usually involves multiple scenarios, where $\mathcal{P}$ is a finite set of more than one probability measures; see, e.g., Kou et al. (2013) and Kou & Peng (2016).

Denote $\mathcal{X}$, a subset of $\mathcal{L}^0(\Omega, \mathcal{F}, \mathcal{P})$, the set of firms' capital positions or portfolio P&Ls at the end of a given period under the risk manager's investigation. Here, the capital position of a firm at the end of a period refers to the firm's asset value minus its liability, and the P&L of a portfolio refers to its profit and loss—the portfolio value at the end of the period minus its initial value. Then, for any $X \in \mathcal{X}$, $L = -X$ denotes the corresponding loss random variable. Examples of $\mathcal{X}$ include $\mathcal{L}^\infty(\Omega, \mathcal{F}, \mathcal{P})$ (the space of all bounded random variables), $\mathcal{L}^p(\Omega, \mathcal{F}, \mathcal{P})$ for some $p \in [1, +\infty)$ (the space of random variables whose L^p norm under any $\mathbb{P} \in \mathcal{P}$ is finite), the space of random variables that are bounded from below, spaces known as Orlicz hearts, and $\mathcal{L}^0(\Omega, \mathcal{F}, \mathcal{P})$ (see, e.g., Delbaen 2002, Frittelli & Gianin 2002, Cheridito & Li 2009, He et al. 2015).

A risk measure and an acceptance set can be mutually representable. Given a risk measure ρ, the acceptance set associated with ρ is defined as $\mathcal{A}_\rho := \{X \mid \rho(X) \leq 0\}$. Conversely, given an acceptance set $\mathcal{A}$, the risk measure associated with the acceptance set is defined as $\rho_\mathcal{A} := \inf\{m \mid X + m \in \mathcal{A}\}$.

2.1. Risk Measures: Axioms and Examples

A risk measure ρ is a mapping from $\mathcal{X}$ to $\mathbb{R}$.[1] The following are some commonly used axioms in the literature for risk measures:

1. Monotonicity: $\rho(X) \leq \rho(Y)$ for any $X, Y \in \mathcal{X}$ such that $X \geq Y$.
2. Strict monotonicity: $\rho(X) \leq \rho(Y)$ for any $X, Y \in \mathcal{X}$ such that $X \geq Y$ and the inequality becomes strict if $X \neq Y$.
3. Positive homogeneity: $\rho(\lambda X) = \lambda\rho(X)$ for any $\lambda > 0$ and $X \in \mathcal{X}$ such that $\lambda X \in \mathcal{X}$.
4. Subadditivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$ for any $X, Y \in \mathcal{X}$ such that $X + Y \in \mathcal{X}$.
5. Quasi-convexity: $\rho(\alpha X + (1 - \alpha)Y) \leq \max(\rho(X), \rho(Y))$ for any $X, Y \in \mathcal{X}$ and $\alpha \in (0, 1)$ such that $\alpha X + (1 - \alpha)Y \in \mathcal{X}$.
6. Convexity: $\rho(\alpha X + (1 - \alpha)Y) \leq \alpha\rho(X) + (1 - \alpha)\rho(Y)$ for any $X, Y \in \mathcal{X}$ and $\alpha \in (0, 1)$ such that $\alpha X + (1 - \alpha)Y \in \mathcal{X}$.
7. Translation invariance with respect to $R \in \mathcal{L}^0(\Omega, \mathcal{F}, \mathcal{P})$, where $R \geq 0$: $\rho(X + mR) = \rho(X) - m$ for any $X \in \mathcal{X}$ and $m \in \mathbb{R}$ such that $X + mR \in \mathcal{X}$.
8. Law invariance: $\rho(X) = \rho(Y)$ for any $X, Y \in \mathcal{X}$ such that they are identically distributed.
9. Comonotonic additivity: $\rho(X + Y) = \rho(X) + \rho(Y)$ for any $X, Y \in \mathcal{X}$ such that they are comonotonic. [X and Y are comonotonic if $(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0$ holds almost surely for any ω_1 and ω_2 in Ω.]
10. Continuity (*a*): $\lim_{d \to 0}\rho(-(X - d)^+) = \rho(-X^+)$, $\lim_{d \to \infty}\rho(-\min(X, d)) = \rho(-X)$, and $\lim_{d \to -\infty}\rho(-\max(X, d)) = \rho(-X)$ hold for any $X \in \mathcal{X}$ and any $x \in \mathbb{R}$, where $x^+ := \max(x, 0)$.
11. Continuity (*b*): $\lim_{d \to \infty}\rho(-\min(\max(X, -d), d)) = \rho(-X)$ for any $X \in \mathcal{X}$.
12. Scale normalization: $\rho(-1) = 1$.
13. Standardization: $\rho(-x) = sx$ for any $x \in \mathbb{R}$, where $s > 0$ is a constant not depending on x.
14. Comonotonic subadditivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$ for any $X, Y \in \mathcal{X}$ such that they are comonotonic.

[1] In some studies, it is assumed that a risk measure can also take $+\infty$ or $-\infty$ (see, e.g., Delbaen 2002, Ruszczyński & Shapiro 2006, He et al. 2015).

15. Comonotonic convexity: $\rho(\lambda X + (1-\lambda)Y) \le \lambda\rho(X) + (1-\lambda)\rho(Y)$ for any $\lambda \in [0, 1]$ and any $X, Y \in \mathcal{X}$ such that they are comonotonic.
16. Comonotonic independence: For any pairwise comonotonic random variables $X, Y, Z \in \mathcal{X}$ and for any $\alpha \in (0, 1)$, $\rho(X) < \rho(Y)$ implies that $\rho(\alpha X + (1-\alpha)Z) < \rho(\alpha Y + (1-\alpha)Z)$.
17. Cash loss: $\rho(-x) = x$ for any $x \in \mathbb{R}_+$.
18. Loss dependence: $\rho(X) = \rho(\min(X, 0))$ for any $X \in \mathcal{X}$.
19. Excess invariance: $\rho(X) = \rho(Y)$ for any $X, Y \in \mathcal{X}$ such that $X^- = Y^-$.
20. Prudence: $\lim_{k\to\infty}\rho(\xi_k) \ge \rho(X)$ holds if $\xi_i \to X$ and $\lim_{k\to\infty}\rho(\xi_k)$ exists.
21. No reward for concentration: There exists an event A such that $\rho(X+Y) = \rho(X) + \rho(Y)$ holds for any X and Y sharing the tail event A.

The (strict) monotonicity axiom and the translation invariance (with respect to R) axiom are commonly assumed in the literature of risk measures. While the former is reasonable and natural, the latter is valid only when the risk measure has a physical meaning. For instance, consider a clearing house that wants to compute the margin requirement for each of its clearing members. Suppose X represents the P&L of a member's portfolio at the end of a given period plus the total return of the member's existing deposit in his margin account, which generates gross return R in the same period (e.g., the risk-free gross return). Thus, $\rho(X)$ refers to the additional deposit the member needs to post in the margin account, where a negative $\rho(X)$ means that the member can take $-\rho(X)$ amount of deposit out of the margin account. If the member increases m dollars in the margin account at the beginning of the period, then the remaining amount the member needs to post, namely $\rho(X) - m$, should be the same as $\rho(X + mR)$, which is the required amount should the clearing house reevaluate the member's position. Thus, translation invariance needs to hold.

The subadditivity axiom is based on the intuition that "a merger does not create extra risk" (Artzner et al. 1999, p. 209), which may not be true. We discuss the controversies related to this axiom in Section 3.

Artzner et al. (1999) propose the coherent risk measures that satisfy the translation invariance, monotonicity, positive homogeneity, and subadditivity axioms. The representation of coherent risk measures for a finite space Ω are obtained by Artzner et al. (1999) based on the results of Huber (1981), who use the same set of axioms. Gilboa & Schmeidler (1989) obtain a more general representation based on a different set of axioms. Delbaen (2002) obtains the representation of coherent risk measures for general probability space, which is given by $\rho(X) = \sup_{Q\in\mathcal{Q}}\{\mathrm{E}^Q[-X]\}$, $\forall X \in \mathcal{X}$, where $\mathcal{Q}$ is a family of probability measures, and $\mathrm{E}^Q[-X]$ is the expectation of $-X$ under the probability measure Q. Kusuoka (2001) derives the representation of law-invariant coherent risk measures. Föllmer & Schied (2002) and Frittelli & Gianin (2002) propose and obtain representations of the convex risk measures that satisfy translation invariance, monotonicity, and convexity axioms, which relax the positive homogeneity and subadditivity axioms. Weber et al. (2013) show that coherent risk measures applied to systems that exhibit price impact may induce convex risk measures that are not positively homogeneous.

Comonotonic random variables are studied by Yaari (1987), Schmeidler (1989), Denneberg (1994), and others. If two random variables X and Y are comonotonic, $X(\omega)$ and $Y(\omega)$ always move in the same direction however the state ω changes. For example, the payoffs of a call option and its underlying asset are comonotonic.

Wang et al. (1997) propose the distortion risk measures (also called insurance risk measures) that satisfy the axioms of law invariance, comonotonic additivity, continuity (*a*), and scale normalization. Wang et al. (1997) show that ρ is a distortion risk measure if and only if ρ has a Choquet integral representation with respect to a distorted probability:

$$\rho(X) = \int (-X)\mathrm{d}(h \circ \mathrm{P}) = \int_{-\infty}^{0} (h(\mathrm{P}(-X > t)) - 1)\mathrm{d}t + \int_{0}^{\infty} h(\mathrm{P}(-X > t))\mathrm{d}t, \qquad 1.$$

where $h(\cdot)$ is a distortion function which is nondecreasing and satisfies $h(0) = 0$ and $h(1) = 1$; $h(\cdot)$ need not be left or right continuous. $h \circ \mathrm{P}$ is called the distorted probability and is defined by $h \circ \mathrm{P}(A) := h(\mathrm{P}(A))$ for any event A. In general, a distortion risk measure does not satisfy subadditivity unless $h(\cdot)$ is concave (Denneberg 1994).

The standardization axiom with $s = 1$ is proposed by Schmeidler (1986); the constant s in the axiom can be related to the countercyclical indexing risk measures proposed by Gordy & 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 & Pedersen 2009, Brunnermeier et al. 2009, Adrian & Shin 2014).

The comonotonic subadditivity axiom is proposed by Kou et al. (2006, 2013) and independently by Song & Yan (2006). Kou et al. (2006, 2013) provide representation of risk statistics (empirical risk measures) that satisfy comonotonic subadditivity and other axioms. Song & Yan (2006) provide representations of risk measures that satisfy positive homogeneity, monotonicity, translation invariance, and comonotonic subadditivity or comonotonic convexity. Song & Yan (2009) give a representation of risk measures that respect stochastic orders and are comonotonically subadditive or convex.

The comonotonic independence axiom is proposed by Kou & Peng (2016). The axiom is related to the comonotonic independence axiom for the Choquet expected utility risk preference (Schmeidler 1989). The motivation for proposing the comonotonic independence axiom is twofold. First, it 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 the axiom, consider three pairwise comonotonic random variables X, Y, and 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, and 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 the axiom. Kou & Peng (2016) show that a risk measure ρ satisfies the set of axioms of comonotonic independence, monotonicity, standardization with $s = 1$, law invariance, and continuity (*b*) if and only if ρ has the representation in Equation 1; the representation result extends that of Wang et al. (1997) as the set of axioms required by Kou & Peng (2016) are weaker. Many commonly used risk measures are special cases of the distortion risk measures in Equation 1.

Cont et al. (2013) propose the cash loss axiom and the loss dependence axiom; the latter is equivalent to the excess invariance axiom proposed by Staum (2013). The loss dependence (excess invariance) axiom postulates that the risk of a portfolio only depends on its losses, which is a desirable property when the risk measure is used for computing margin requirements or capital reserves. Cont et al. (2013) propose the loss-based risk measures that satisfy the axioms of cash loss, monotonicity, and loss dependence. They provide the representation of convex loss-based risk measures and law-invariant convex loss-based risk measures. Staum (2013) proposes the shortfall risk measures that are normalized [i.e., $\rho(0) = 0$], are nonnegative, and satisfy the axioms of monotonicity and excess invariance.

Example 1 (Value-at-risk). VaR is a quantile of the loss distribution at some predefined probability level. More precisely, let X be the P&L; then, $-X$ is the loss with 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_{-X}—i.e., $\mathrm{VaR}_\alpha(X) := q_\alpha^-(F_{-X}) := F_{-X}^{-1}(\alpha) = \inf\{x \mid F_{-X}(x) \geq \alpha\}$. For $\alpha = 0$, VaR of X at level α is defined to be $\mathrm{VaR}_0(X) := \inf\{x \mid F_{-X}(x) > 0\}$

and $\mathrm{VaR}_0(X)$ is equal to the essential infimum of $-X$. VaR is a distortion risk measure. In fact, for $\alpha \in (0, 1]$, ρ in Equation 1 is equal to VaR_α if $h(x) := 1_{\{x > 1-\alpha\}}$, and ρ in Equation 1 is equal to VaR_0 if $h(x) := 1_{\{x=1\}}$.

VaR does not satisfy the subadditivity axiom in general, so it is not a coherent risk measure. VaR has two other important properties: ordinal covariance and monotonicity with respect to first-order stochastic dominance (see, e.g., Denneberg 1994). Chambers (2009) shows that the two properties essentially characterize VaR. Duffie & Pan (1997, 2001), Gordy (2003), Jorion (2007), and Hull (2009) provide comprehensive discussions of VaR and risk management.

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

$$\mathrm{ES}_\alpha(X) := \int_{-\infty}^{\infty} x \mathrm{d}F_{\alpha,-X}(x),\ \alpha \in [0, 1),$$

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

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

For $\alpha = 1$, ES of X at level α is defined as $\mathrm{ES}_1(X) := F_{-X}^{-1}(1)$. If the loss distribution F_{-X} is continuous, then $F_{\alpha,-X}$ is the same as the conditional distribution of $-X$ given that $-X \geq \mathrm{VaR}_\alpha(X)$; if F_{-X} is not continuous, then $F_{\alpha,-X}(x)$ is a slight modification of the conditional loss distribution. For $\alpha \in [0, 1)$, $\rho(X)$ in Equation 1 is equal to $\mathrm{ES}_\alpha(X)$ if $h(x) = x/(1-\alpha)$, if $x \leq 1-\alpha$, and $h(x) = 1$ otherwise. For $\alpha = 1$, $\rho(X)$ in Equation 1 is equal to $\mathrm{ES}_1(X)$ if $h(x) = 1_{\{x>0\}}$. Tasche (2002, proposition 3.4) shows that

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 F_{-X}^{-1}(s)\mathrm{d}s = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_s(X)\mathrm{d}s. \qquad 2.$$

ES is also called conditional VaR and superquantile. Rockafellar et al. (2014) show that $\mathrm{ES}_\alpha(X) = \arg\min_{c\in\mathbb{R}} \mathcal{E}_\alpha(X - c)$, where $\mathcal{E}_\alpha(\cdot)$ is an error function; based on such representation of $\mathrm{ES}_\alpha(X)$, they develop the superquantile regression for the estimation of ES_α.

Wang & Zitikis (2021) propose the axioms of prudence and no reward for concentration. The former yields the lower semicontinuity of the risk measure; the latter means that if X and Y incur large loss simultaneously in the stress event A, then the capital requirement of the portfolio $X + Y$ does not receive reduction. They show that ES can be characterized by the two axioms along with the axioms of monotonicity and law invariance.

Example 3 (Median shortfall). As we discuss later, ES has several statistical drawbacks, including nonelicitability and nonrobustness. To mitigate the problems, one may simply use MS. 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)[2]

$$\mathrm{MS}_\alpha(X) := \text{median of the } \alpha\text{-tail distribution of the loss} - X = F_{\alpha,-X}^{-1}\left(\frac{1}{2}\right).$$

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 VaR at level α, because

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

it measures the median of the loss size conditional on that the loss exceeds VaR at level α. Kou et al. (2013) show that

$$\mathrm{MS}_\alpha(X) = \mathrm{VaR}_{\frac{1+\alpha}{2}}(X),\ \forall X,\ \forall \alpha \in [0,1]. \tag{3}$$

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

Since $\mathrm{MS}_\alpha = \mathrm{VaR}_{(1+\alpha)/2}$, we know that 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 like MS_α, ES_α can also fail to reveal large loss beyond $\mathrm{VaR}_{(1+\alpha)/2}$. For example, let $\lambda, \mu > 0$, 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

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

where $\beta(n) := \frac{\mu}{n-c-\frac{1}{\lambda}}$. 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 $c + \mu + \frac{1}{\lambda}$; hence, ES_α fails in the same way as MS_α regarding the detection of the large loss with size n that 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

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

where Δ is a probability measure on (0, 1]. Kou & Peng (2016) show that the class of distortion risk measures represented by Equation 1 includes and is strictly larger than the class of generalized spectral risk measures. In fact, 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 Equation 1, with $h(x)$ being defined as $h(x) := 1_{\{x \ge 1-\alpha\}}$, but q_α^+ cannot be represented by Equation 5. A special case of Equation 5 is the spectral risk measure (Acerbi 2002, definition 3.1), defined as

$$\rho(X) = \int_{(0,1)} F_{-X}^{-1}(u)\phi(u)\mathrm{d}u,\ \phi(\cdot) \text{ is increasing, nonnegative, and } \int_0^1 \phi(u)\mathrm{d}u = 1. \tag{6}$$

Because of the requirement that ϕ is increasing, the class of spectral risk measures is much smaller than the class of generalized spectral risk measures defined in Equation 5. The distinction between the spectral risk measure and that in Equation 5 is that the former is convex but the latter may not be convex. The convexity requires that the function ϕ in Equation 6 is an increasing function. The MINMAXVAR risk measure proposed in Cherny & Madan (2009) for the measurement of trading performance is a special case of the spectral risk measure, corresponding to a distortion function $h(x) = 1 - (1 - x^{\frac{1}{1+\alpha}})^{1+\alpha}$ in Equation 1 with $s = 1$, where $\alpha \ge 0$ is a constant.

The class of distortion risk measures and the class of law-invariant coherent (convex) risk measures have nonempty intersections, but one is not the subset of the other. For example, ES belongs to both classes; VaR belongs to the former but not the latter. The class of distortion risk measures includes the class of law-invariant spectral risk measures as a strict subset. For example, VaR belongs to the former but not the latter.

2.2. Acceptance Sets

The regulator can propose a capital adequacy test by specifying an acceptance set $\mathcal{A} \subset \mathcal{X}$ for firms' capital positions at time T; a formal definition of capital positions and the motivation for the test is provided by, for instance, He & Peng (2018). The following axioms for the acceptance set $\mathcal{A}$ have been proposed in the literature.

1. Positive inclusion: $X \in \mathcal{A}$ for any $X \in \mathcal{X}$ such that $X \geq 0$.
2. Strict negative exclusion: $X \notin \mathcal{A}$ for any $X \in \mathcal{X}$ such that $X < 0$.
3. Convexity: If $X, Y \in \mathcal{A}$, then for any $\alpha \in (0, 1)$ such that $\alpha X + (1-\alpha)Y \in \mathcal{X}$, $\alpha X + (1-\alpha)Y \in \mathcal{A}$.
4. Conicity: If $X \in \mathcal{A}$, then for any $\lambda > 0$ such that $\lambda X \in \mathcal{X}$, $\lambda X \in \mathcal{A}$.
5. Monotonicity: If $X \in \mathcal{A}$, then for any $Y \in \mathcal{X}$ such that $Y \geq X$, $Y \in \mathcal{A}$.
6. Law invariance: For any $X, Y \in \mathcal{X}$ such that X and Y are identically distributed, if $X \in \mathcal{A}$, then $Y \in \mathcal{A}$.
7. Numéraire invariance: For any $X \in \mathcal{X}$ and any strictly positive random variable Z on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $ZX \in \mathcal{X}$, if $X \in \mathcal{A}$, then $ZX \in \mathcal{A}$.
8. Surplus invariance: For any $X, Y \in \mathcal{X}$, if $X \in \mathcal{A}$ and $X^- \geq Y^-$ almost surely, then $Y \in \mathcal{A}$.
9. Truncation-closedness: For any $X \in \mathcal{X}$, if $\min(\max(-d, X), d) \in \mathcal{A}$ for any $d > 0$, then $X \in \mathcal{A}$.

The convexity and conicity axioms for an acceptance set correspond to the convexity and positive homogeneity axioms for a risk measure. Conicity stipulates that scaling the capital position of a firm by a positive constant does not change the acceptability of the firm. Artzner et al. (1999) show that a coherent risk measure corresponds to a coherent acceptance set that satisfies the axioms of positive inclusion, strict negative exclusion, convexity, and conicity. Föllmer & Schied (2002) show that a convex risk measure corresponds to a convex acceptance set that satisfies the axioms of convexity and monotonicity.

The numéraire invariance axiom is introduced by Artzner et al. (2009) and further investigated by Koch-Medina et al. (2017). It means that the acceptance set should not depend on the choice of the numéraire asset. When a capital adequacy test is applied internationally, it may be desirable to be numéraire invariant: Whether a firm passes the test should not depend on the currency that is used to denominate the firm's capital position. Koch-Medina & Munari (2016, section 6) show that the acceptance set associated with VaR is numéraire invariant but that associated with ES is not. Clearly, numéraire invariance implies conicity; furthermore, Koch-Medina et al. (2017, proposition 5) show that a closed acceptance set $\mathcal{A}$ is numéraire invariant and monotone if and only if it is surplus invariant and conic. Without assuming numéraire invariance, the currency risk can also be explicitly incorporated in capital adequacy tests by using vector-valued risk measures (see, e.g., Jouini et al. 2004).

The surplus invariance axiom is proposed by Koch-Medina et al. (2015),[3] extending the excess invariance axiom of the shortfall risk measures and the loss dependence axiom of the loss-based risk measures proposed by Staum (2013) and Cont et al. (2013), respectively. X^-, the negative part of a firm's capital position X, is its option to default, and the positive part of X is the surplus of the firm's shareholders. The surplus invariance axiom stipulates that if firm A passes the test, then firm B, whose option to default is smaller than that of firm A, should also pass the test.

[3] Koch-Medina et al. (2015) use a slightly different, but equivalent, definition of surplus invariance (see equation 1.1 in their paper and Koch-Medina et al. 2017, proposition 1).

Staum (2013) shows that the surplus invariance axiom is satisfied by the acceptance set that is associated with VaR, i.e., $\mathcal{A} = \{X \in \mathcal{X} \mid \mathrm{VaR}_\alpha(X) \leq 0\}$, and the one that is associated with a shortfall risk measure, i.e., $\mathcal{A} = \{X \in \mathcal{X} \mid \mathbb{E}[l(X^-)] \leq c\}$, where l is a nonconstant and increasing function. In contrast, Koch-Medina et al. (2015) find that the acceptance set associated with ES, i.e., $\mathcal{A} = \{X \in \mathcal{X} \mid \mathrm{ES}_\alpha(X) \leq 0\}$, is not surplus invariant. Koch-Medina et al. (2015, 2017) provide dual characterizations of convex and surplus-invariant acceptance sets. Furthermore, they prove that the only coherent acceptance set that is simultaneously law invariant and surplus invariant is the positive cone $L_+ = \{X \in \mathcal{X} \mid X \geq 0\}$.

The truncation-closedness axiom means that if any truncated version of the (possibly unbounded) capital position X is acceptable, then X itself is also acceptable. This axiom is similar to the continuity axiom for the distortion risk measure ρ (Kou & Peng 2016). The truncation-closedness axiom automatically holds for any acceptance set if $\mathcal{X} = \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P})$.

He & Peng (2018) show that surplus-invariant, law-invariant, conic (or numéraire-invariant), and truncation-closed acceptance sets must be the sets induced by VaR at some confidence level α, and such acceptance sets cannot be induced by ES. The result highlights the relevance of using VaR in capital adequacy tests. Because the surplus invariance and law invariance axioms are indispensable to a large extent, to use other acceptance sets, such as convex ones, one has to drop the conicity axiom and thus the numéraire invariance axiom. This, however, means that capital adequacy tests based on those acceptance sets are sensitive to the choice of denominating currencies.

3. CONTROVERSY OF THE SUBADDITIVITY AND CONVEXITY AXIOMS

Although the subadditivity and convexity axioms yield some mathematical convenience, we give some economic and statistical reasons to relax them from the following aspects: (*a*) diversification may not always be beneficial, (*b*) a merger may create extra risk, and (*c*) other issues of subadditivity may apply.

3.1. Diversification May Not Always Be Beneficial

Both the subadditivity and convexity axioms are related to the idea that diversification does not increase risk. There are two main justifications for diversification. One is based on the simple observation that $\sigma(X+Y) \leq \sigma(X) + \sigma(Y)$, for any two random variables X and Y with finite second moments, where $\sigma(\cdot)$ denotes standard deviation. The other is based on expected utility theory. Samuelson (1967) shows that any investor with a strictly concave utility function will uniformly diversify among independently and identically distributed (i.i.d.) risks with finite second moments [see, e.g., McMinn (1984), Hong & Herk (1996), and Kijima (1997) for the discussion on whether diversification is beneficial when the asset returns are dependent]. Both justifications require that the risks have finite second moments.

Is diversification still preferable for risks with infinite second moments? The answer can be no. Fama & Miller (1972, pp. 271–72) show that diversification is ineffective for asset returns with heavy tails (with tail index less than 1). Ibragimov (2004, 2009) and Ibragimov & Walden (2007) show that diversification is not preferable for risks with extremely heavy-tailed distributions (with tail index less than 1) in the sense that (*a*) the loss of the diversified portfolio stochastically dominates that of the undiversified portfolio at the first order and second order, and (*b*) the expected utility of the (truncated) payoff of the diversified portfolio is smaller than that of the undiversified portfolio. They also show that investors with certain S-shaped utility functions would prefer nondiversification, even for bounded risks. The possibility of S-shaped utility functions is supported by experimental results and prospect theory (Kahneman & Tversky 1979, Tversky & Kahneman 1992).

In addition, the conclusion that VaR prohibits diversification, drawn from simple examples in the literature, may not be solid, for the following reasons. First, Artzner et al. (1999, pp. 217–18) show that VaR prohibits diversification by a simple example in which 95% VaR of the diversified portfolio is higher than that of the undiversified portfolio. However, in the same example, 99% VaR encourages diversification because the 99% VaR of the diversified portfolio is equal to 20,800, which is much lower than 1,000,000, the 99% VaR of the undiversified portfolio. Second, McNeil et al. (2005, p. 241, example 6.7) show that for $\alpha = 95\%$, the VaR_α of a fully concentrated portfolio is smaller than that of a fully diversified portfolio. However, for $\alpha > 98\%$, such a pathological example no longer holds.

Even if one believes in subadditivity, VaR and MS satisfy subadditivity in most relevant situations. In fact, Daníelsson et al. (2013) show that VaR and MS are subadditive in the relevant tail region if asset returns are regularly varying and possibly dependent, although VaR does not satisfy global subadditivity. Ibragimov (2004, 2009) and Ibragimov & Walden (2007) also show that although VaR does not satisfy subadditivity for risks with extremely heavy-tailed distributions (with tail index less than 1), VaR satisfies subadditivity for wide classes of independent and dependent risks with tail indices greater than 1, such as the infinite variance stable distributions with finite mean. In the words of Gaglianone et al. (2011, p. 150), "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)." 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."

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

3.2. A Merger May Create Extra Risk

Subadditivity basically means that "a merger does not create extra risk" (Artzner et al. 1999, p. 209). However, Dhaene et al. (2003) point out that a merger may increase risk, particularly when there is bankruptcy protection for institutions. For example, an institution can split a risky trading business into a separate subsidiary so that it has the option to let the subsidiary go bankrupt when the subsidiary suffers enormous losses, confining losses to that subsidiary. Therefore, creating subsidiaries may incur less risk, and a merger may increase risk. Had Barings Bank set up a separate institution for its Singapore unit, the bankruptcy of that unit would not have sunk the entire bank in 1995.

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

3.3. Other Issues of Subadditivity May Apply

Subadditivity is not necessarily needed for capital allocation or asset allocation. 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; Shi & Werker (2012), Wen et al. (2013), Xi et al. (2014), and the references therein provide asset allocation models and methods based on VaR and Basel Accord risk measures.

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 a firm has an incentive to merge with other firms, which may lead to financial firms 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.

Although subadditivity of ρ ensures that $\rho(X_1) + \rho(X_2)$ is an upper bound for $\rho(X_1 + X_2)$, this upper bound may not be valid in face of model uncertainty. In fact, suppose we are concerned with obtaining an upper bound for $\mathrm{ES}_\alpha(X_1 + X_2)$. In practice, due to model uncertainty, we can only compute $\widehat{\mathrm{ES}}_\alpha(X_1)$ and $\widehat{\mathrm{ES}}_\alpha(X_2)$, which are estimates of $\mathrm{ES}_\alpha(X_1)$ and $\mathrm{ES}_\alpha(X_2)$ respectively. $\widehat{\mathrm{ES}}_\alpha(X_1) + \widehat{\mathrm{ES}}_\alpha(X_2)$ cannot be used as an upper bound for $\mathrm{ES}_\alpha(X_1 + X_2)$ because it is possible that $\widehat{\mathrm{ES}}_\alpha(X_1) + \widehat{\mathrm{ES}}_\alpha(X_2) < \mathrm{ES}_\alpha(X_1) + \mathrm{ES}_\alpha(X_2)$ due to model misspecification.

In addition, $\rho(X_1) + \rho(X_2)$ may not be a useful upper bound for $\rho(X_1 + X_2)$ as the former may be too much larger than the latter. For example, let X_1 be the P&L of a long position of a call option on a stock (whose price is \$100) at strike \$100 and let X_2 be the P&L 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 $-(X_1 + X_2) \leq 5$. However, $\rho(X_1) = 0$ and $\rho(X_2) \approx 20$ (the margin is about 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 $\rho(X_1 + X_2)$.

Finally, the theory of individual choice, particularly prospect theory, suggests that it may be appropriate to relax the subadditivity, which motivates the postulation of the comonotonic subadditivity axiom. There are simple examples showing that the risk associated with noncomonotonic random variables can violate subadditivity because people are risk seeking instead of risk averse when facing losses of medium or large probabilities, as implied by prospect theory. In fact, consider the following simple example: Suppose there is an urn that contains 50 black balls and 50 red balls. Randomly draw a ball from the urn. Let B be the position of losing \$10,000 in the event that the ball is black, and let R be the position of losing \$10,000 in the event that the ball is red. Obviously, B and R bear the same amount of risk, i.e., $\rho(B) = \rho(R)$. Let S be the event of losing \$5,000 for sure; then, $\rho(S) = 5{,}000$. According to prospect theory, people are risk seeking in choices between probable and sure losses—i.e., most people would prefer a larger loss with a substantial probability to a sure loss. Most people would thus prefer position B to position S (see Kahneman & Tversky 1979, p. 273, problem 12; Tversky & Kahneman 1992, p. 307, table 3). In other words, $\rho(B) = \rho(R) < \rho(S) = 5{,}000$. Since the position $B + R$ corresponds to a sure loss of \$10,000, $\rho(B + R) = 10{,}000$. Combining them, $\rho(B) + \rho(R) < 5{,}000 + 5{,}000 = 10{,}000 = \rho(B + R)$, which violates subadditivity. Clearly the random losses associated with B and R are not comonotonic. In addition, even in terms of expected utility theory, it is not clear whether a risk measure should be superadditive or subadditive for independent random variables (see Eeckhoudt & Schlesinger 2006).

4. ROBUSTNESS OF RISK MEASURES

A risk measure is said to be robust if (*a*) it is insensitive to small changes in the data—i.e., a small change of the data set, such as changing a few samples, or adding a few outliers to the data set, or making small changes to many samples, only results in a small change to the estimated risk measure—and (*b*) it can accommodate model misspecification (possibly by incorporating multiple scenarios and models).

The first part of the meaning of robustness comes from the study of robust statistics, which is mainly concerned with the statistical (distributional) robustness (see, e.g., Huber & Ronchetti 2009). The second part of the meaning of robustness is related to ambiguity and model uncertainty in decision theory. To address these issues, multiple priors or multiple alternative models represented by a set of probability measures may be used (see, e.g., Gilboa & Schmeidler 1989, Maccheroni et al. 2006, Hansen & Sargent 2007).

4.1. Robustness is Indispensable for External Risk Measures

Kou et al. (2006, 2013) argue that an external risk measure used for financial regulation should be robust, because (*a*) legal realism, a basic concept in law, emphasizes that robustness is essential for law enforcement, and (*b*) for consistent implementation of an external risk measure across all relevant financial institutions, the risk measure needs to be robust with respect to noise in the data and unavoidable model misspecification.

Legal realism is the viewpoint that the legal decisions of a court are determined by the actual practices of the judges rather than the law set forth in statutes and precedents. All the legal rules contained in statutes and precedents are uncertain because of the uncertainty in human language and because human beings are unable to anticipate all possible future circumstances (Hart 1994, p. 128). Hence, a law is only a guideline for judges and enforcement officers (Hart 1994, pp. 204–5)—that is, it is only intended to be the average of what judges and officers will decide. This concerns the robustness of law; i.e., a law should be established in such a way that different judges will reach similar conclusions when they implement it. In particular, consistent enforcement of an external risk measure in banking regulation requires that it be robust with respect to underlying models and data.

In determining capital requirements, regulators impose a risk measure and allow institutions to use their own internal risk models and private data in the calculation. However, there are two issues arising from the use of internal models and private data in external regulation: (*a*) The data can be noisy, flawed, or unreliable, and (*b*) there can be several statistically indistinguishable models for the same asset or portfolio because of limited availability of data. For example, the heaviness of tail distributions cannot be identified in many cases. Heyde & Kou (2004) show that it is very difficult to distinguish between exponential-type and power-type tails with 5,000 observations (about 20 years of daily observations) because the estimated quantiles of exponential-type distributions and power-type distributions may overlap.

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

4.2. Statistical Robustness of Value-at-Risk and Median Shortfall

VaR and MS are shown to be more robust than ES by four tools in robust statistics: influence functions, asymptotic breakdown points, finite sample breakdown points, and Hampel robustness. We now describe these in more detail.

1. Let F be the distribution function of X. Let $X_1, X_2, \ldots$ be a sequence of independent observations with distribution F. Let $F_n(x) := \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le x\}}$ be the empirical distribution. The influence function of a statistical functional T at F and a point $y \in \mathbb{R}$ is defined as $IF(y, T, F) := \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon}\left[T((1-\varepsilon)F + \varepsilon\delta_y) - T(F)\right]$, $y \in \mathbb{R}$, where δ_y is the point mass 1 at y that represents a contamination point to the distribution F. If the influence function is bounded, i.e., $\sup_{y\in\mathbb{R}} |IF(y, T, F)| < \infty$, then $T(F_n)$ is robust; otherwise, $T(F_n)$ is not robust, and outliers in the data may cause large changes to $T(F_n)$ (Huber & Ronchetti 2009). Kou et al. (2006, 2013) show that VaR and MS have bounded influence functions but ES has an unbounded one.
2. The asymptotic breakdown point is, roughly, the smallest fraction of bad observations that may cause an estimator to take on arbitrarily large aberrant values (for the mathematical definition, see Huber & Ronchetti 2009, section 1.4). Hence, a high breakdown point is clearly desirable. Kou et al. (2006, 2013) show that the asymptotic breakdown point of VaR_α (resp. MS_α) is $1-\alpha$ [resp. $(1-\alpha)/2$] and that of ES_α is 0.
3. The finite sample breakdown point (Huber & Ronchetti 2009, chapter 11) of $\mathrm{VaR}_\alpha(F_n)$ [resp. $\mathrm{MS}_\alpha(F_n)$] is $(n - \lceil n\alpha\rceil + 1)/(2n - \lceil n\alpha\rceil + 1) \approx (1-\alpha)/(2-\alpha)$ [resp. $(n - \lceil n(1+\alpha)/2\rceil + 1)/(2n - \lceil n(1+\alpha)/2\rceil + 1) \approx (1-\alpha)/(3-\alpha)$], but that of $\mathrm{ES}_\alpha(F_n)$ is $1/(n+1)$, which means one additional corrupted sample can cause arbitrarily large bias to ES_α.
4. Hampel robustness (Hampel 1971) is one commonly accepted definition of distributional robustness. Let $T_n = T_n(X_1, \ldots, X_n)$ be a sequence of estimates. Let $\mathcal{L}_F(T_n)$ be the distribution of T_n under F. The sequence of estimates T_n is called Hampel robust at F_0 if the sequence of maps $F \to \mathcal{L}_F(T_n)$ is equicontinuous at F_0—i.e., if for any $\epsilon > 0$, there exists $\delta > 0$ and $n_0 > 0$, such that for all F and all $n \ge n_0$, $d_*(F, F_0) < \delta \Rightarrow d_*(\mathcal{L}_F(T_n), \mathcal{L}_{F_0}(T_n)) < \epsilon$, where d_* is any metric generating the weak topology, such as the Prokhorov metric and Lévy metric. Let $q_\alpha^+(F) := \inf\{x : F(x) > \alpha\}$ be the right α-quantile of F. Cont et al. (2010) show that (*a*) for any α and any F_0 such that $q_\alpha^-(F_0) = q_\alpha^+(F_0)$ [resp. $q^-_{(1+\alpha)/2}(F_0) = q^+_{(1+\alpha)/2}(F_0)$], $\mathrm{VaR}_\alpha(F_n)$ [resp. $\mathrm{MS}_\alpha(F_n)$] is Hampel robust at F_0, and (*b*) for any α and any F_0, $\mathrm{ES}_\alpha(F_n)$ is not Hampel robust at F_0.

Krätschmer et al. (2014) develop a notion of qualitative robustness that is more refined than Hampel robustness. The degree of robustness of a risk measure under this notion can be expressed by an index of qualitative robustness that takes values in $[0, \infty]$, with the value ∞ corresponding to Hampel robustness. Under this notion of robustness, VaR is still more robust than ES and expectile, as the index of qualitative robustness of VaR is ∞ but those of ES and expectile are both equal to 1.

Embrechts et al. (2015) introduce the notion of aggregation robustness, which is a notion of robustness that is weaker than Hampel robustness. By theorem 2.21 of Huber & Ronchetti (2009), saying that a risk measure (statistical functional) ρ is Hampel robust at a distribution F is essentially equivalent to saying 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, if ρ is aggregation robust at F, it 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. Embrechts et al. (2015) argue that with respect to dependence uncertainty in aggregation, VaR is less robust compared to expected shortfall, because VaR is not aggregation robust but ES is. However, their counterexample (Embrechts et al. 2015, 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 $\alpha \in (0, 1)$; in fact, if F is a continuous distribution, then no such α exists. On the contrary, for any other α, VaR at level α is aggregation robust, because VaR at level α is Hampel robust and Hampel robustness implies aggregation robustness; note that by Cont et al. (2010, corollary 3.7), ES is not Hampel robust.

ES is also highly model dependent and particularly sensitive to modeling assumptions on the extreme tails of loss distributions because the computation of ES relies on these extreme tails, as is shown by Equation 2. Kou et al. (2013, figure 1) illustrate the sensitivity of ES to modeling assumptions; MS is less sensitive to tail behavior than ES because the changes of MS with respect to the changes of loss distributions have narrower ranges than do those of ES.

MS is statistically robust as it uses the median of the tail distribution and the median is robust. Other robust measures of the tail distribution, such as the trimmed mean, may also be used to generate robust risk measures. Cont et al. (2010) propose the range value-at-risk (RVaR) measure $\rho(X) = \frac{1}{\alpha_2 - \alpha_1} \int_{\alpha_1}^{\alpha_2} F_{-X}^{-1}(u) \mathrm{d}u$, where $0 < \alpha_1 < \alpha_2 < 1$. RVaR is Hampel robust; however, RVaR is not elicitable. In fact, RVaR is a special case of spectral risk measure and hence a special case of the distortion risk measure defined in Equation 1. Kou & Peng (2016) show that the only two elicitable risk measures among the class of distribution risk measures are VaR (including MS) and the mean functional.

4.3. Robust Risk Measures Incorporating Multiple Scenarios (Prior Probability Measures)

Kou et al. (2006, 2013) propose the natural risk statistics, a class of data-based risk measures that incorporate multiple prior probability measures (scenarios) and robust risk measurement under each scenario. In external regulation, the behavior of the random variable X under different scenarios is preferably represented by different sets of data observed or generated under those scenarios because specifying accurate models for X (under different scenarios) is usually very difficult. More precisely, suppose the behavior of X is represented by a collection of data $\tilde{x} = (\tilde{x}^1, \tilde{x}^2, \ldots, \tilde{x}^m) \in \mathbb{R}^n$, where $\tilde{x}^i = (x_1^i, \ldots, x_{n_i}^i) \in \mathbb{R}^{n_i}$ is the data subset that corresponds to the ith scenario and n_i is the sample size of $\tilde{x}^i$, $n_1 + n_2 + \cdots + n_m = n$. For each $i = 1, \ldots, m$, $\tilde{x}^i$ can be a data set based on historical observations, hypothetical samples simulated according to a model, or a mixture of observations and simulated samples. X can be either discrete or continuous. For example, the data used in the calculation of the Basel III risk measure comprise 120 data subsets corresponding to 120 different scenarios ($m = 120$); Section 7 discusses the details of the Basel III risk measures.

A risk statistic $\hat{\rho}$ is simply a mapping from $\mathbb{R}^n$ to $\mathbb{R}$. The advantage of using risk statistics include the following: (*a*) risk statistics can directly measure risk from observations without specifying subjective models, which greatly reduces model misspecification error; (*b*) risk statistics can incorporate forward-looking views or prior knowledge by including data subsets generated by models based on such views or knowledge; and (*c*) risk statistics can incorporate multiple prior probabilities on the set of scenarios that reflect multiple beliefs about the probabilities of occurrence of different scenarios.

Let $\tilde{x} = (\tilde{x}^1, \tilde{x}^2, \ldots, \tilde{x}^m) \in \mathbb{R}^n$ and $\tilde{y} = (\tilde{y}^1, \tilde{y}^2, \ldots, \tilde{y}^m) \in \mathbb{R}^n$ represent the observations of random losses X and Y under m scenarios. The notion of scenario-wise comonotonicity for $\tilde{x}$ and $\tilde{y}$ is the counterpart of the notion of comonotonicity for the two random variables X and Y. $\tilde{x}$ and $\tilde{y}$ are said to be scenario-wise comonotonic if for $\forall i$, $\forall 1 \leq j, k \leq n_i$, it holds that $(x^i_j - x^i_k)(y^i_j - y^i_k) \geq 0$. Kou et al. (2006, 2013) propose the following axiom for a risk statistic $\hat{\rho}$:

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

The axiom relaxes the subadditivity requirement in coherent risk measures and relaxes the comonotonic additivity requirement in distortion risk measures. A risk statistic is called a natural risk statistic if it satisfies the axioms of monotonicity, positive homogeneity, translation invariance with respect to $R = 1/s$ ($s > 0$ is a constant), scenario-wise comonotonic subadditivity, and empirical law invariance. Kou et al. (2006, 2013) show that $\hat{\rho}$ is a natural risk statistic if and only if $\hat{\rho}$ is represented by

$$\hat{\rho}(\tilde{x}) := s \cdot \sup_{\tilde{w} \in \mathcal{W}} \left\{ \sum_{j=1}^{n_1} w^1_j l^1_{(j)} + \sum_{j=1}^{n_2} w^2_j l^2_{(j)} + \cdots + \sum_{j=1}^{n_m} w^m_j l^m_{(j)} \right\}, \ \forall \tilde{x} = (\tilde{x}^1, \ldots, \tilde{x}^m) \in \mathbb{R}^n, \qquad 7.$$

where $\tilde{l} = -\tilde{x}$, $\tilde{l} = (\tilde{l}^1, \ldots, \tilde{l}^m)$, and $(l^i_{(1)}, \ldots, l^i_{(n_i)})$ is the order statistics of $\tilde{l}^i = (l^i_1, \ldots, l^i_{n_i})$, with $l^i_{(n_i)}$ being the largest, $i = 1, \ldots, m$; $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^n$ is a set of weights; and each weight $\tilde{w} = (w^1_1, \ldots, w^1_{n_1}, \ldots, w^m_1, \ldots, w^m_{n_m}) \in \mathcal{W}$ satisfies

$$\sum_{j=1}^{n_1} w^1_j + \sum_{j=1}^{n_2} w^2_j + \cdots + \sum_{j=1}^{n_m} w^m_j = 1, \ w^i_j \geq 0, j = 1, \ldots, n_i; i = 1, \ldots, m. \qquad 8.$$

Natural risk statistics include subclasses that are robust. Let $\hat{\rho}$ be a natural risk statistic defined in Equation 7 that corresponds to the set of weights $\mathcal{W}$. Define the map $\phi : \mathcal{W} \to \mathbb{R}^m \times \mathbb{R}^n$ such that $\tilde{w} \mapsto \phi(\tilde{w}) := (\tilde{p}, \tilde{q})$, where $\tilde{p} := (p^1, \ldots, p^m)$, $p^i := \sum_{j=1}^{n_i} w^i_j$, $i = 1, \ldots, m$; $\tilde{q} := (q^1_1, \ldots, q^1_{n_1}, \ldots, q^m_1, \ldots, q^m_{n_m})$, $q^i_j := 1_{\{p^i > 0\}} w^i_j / p^i$. Since $p^i \geq 0$ and $\sum_{i=1}^m p^i = 1$, $\tilde{p}$ can be viewed as a prior probability distribution on the set of scenarios. Then $\hat{\rho}$ can be rewritten as

$$\hat{\rho}(\tilde{x}) = s \cdot \sup_{(\tilde{p}, \tilde{q}) \in \phi(\mathcal{W})} \left\{ \sum_{i=1}^m p^i \hat{\rho}^{i,\tilde{q}}(\tilde{x}^i) \right\}, \quad \text{where} \quad \hat{\rho}^{i,\tilde{q}}(\tilde{x}^i) := \sum_{j=1}^{n_i} q^i_j l^i_{(j)}. \qquad 9.$$

Each weight $\tilde{w} \in \mathcal{W}$ then corresponds to $\phi(\tilde{w}) = (\tilde{p}, \tilde{q}) \in \phi(\mathcal{W})$, which specifies: (*a*) the prior probability measure $\tilde{p}$ on the set of scenarios and (*b*) the subsidiary risk statistic $\hat{\rho}^{i,\tilde{q}}$ for each scenario i, $i = 1, \ldots, m$. Hence, $\hat{\rho}$ can be robust with respect to model misspecification by incorporating multiple prior probabilities $\tilde{p}$ and multiple risk statistics $\hat{\rho}^{i,\tilde{q}}$ for each scenario. In addition, $\hat{\rho}$ can be robust with respect to small changes in the data if each subsidiary risk statistic $\hat{\rho}^{i,\tilde{q}}$ is a robust statistic, such as VaR or MS. In particular, VaR and MS with scenario analysis, such as the Basel II and Basel III risk measures (see their definition in Section 7), are robust natural risk statistics.

In contrast, Kou et al. (2006, 2013) show that no law-invariant coherent risk measure is robust with respect to small changes in the data. In fact, they show that a risk statistic $\hat{\rho}$ is an empirical-law-invariant coherent risk statistic if and only if $\hat{\rho}$ is in the form of Equation 7, with the additional constraint that each weight $\tilde{w} \in \mathcal{W}$ must satisfy $w^i_1 \leq w^i_2 \leq \cdots \leq w^i_{n_i}$, $i = 1, \ldots, m$. Hence, any empirical-law-invariant coherent risk statistic assigns larger weights to larger loss observations because both $l^i_{(j)}$ and w^i_j increase when j increases, but assigning larger weights to larger loss observations is clearly sensitive to small changes in the data. Indeed, the finite sample breakdown point (for definition, see, e.g., Huber & Ronchetti 2009, chapter 11) of any empirical-law-invariant coherent risk statistic is equal to $1/(1 + n)$, which implies that one single contamination sample can cause unbounded bias.

Kou & Peng (2016) propose a class of multiple-scenario-based distortion risk measures that include robust subclasses. 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 or a financial crisis. Suppose that under the ith scenario, the measurement of risk is given by a distortion risk measure $\rho_i(X) = \int(-X)\mathrm{d}(h_i \circ P_i)$, where h_i is a distortion function, $i = 1, \ldots, m$. Kou & Peng (2016) propose the following risk measure to incorporate multiple scenarios:

$$\rho(X) = f(\rho_1(X), \rho_2(X), \ldots, \rho_m(X)), \tag{10.}$$

where $f : \mathbb{R}^m \to \mathbb{R}$ is called a scenario aggregation function. Consider the following axiom for the scenario aggregation function f:

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

The axiom is proposed by Gilboa & Schmeidler (1989) to capture the phenomenon of hedging; it is used as one of the axioms for the maxmin expected utility that incorporates robustness. Kou & Peng (2016) show that the scenario aggregation function f in Equation 10 satisfies the axioms of positive homogeneity, translation invariance, monotonicity, and uncertainty aversion if and only if $\rho(X)$ in Equation 10 is represented by[4]

$$\rho(X) = s \cdot \sup_{\tilde{w}\in\mathcal{W}} \left\{ \sum_{i=1}^{m} w_i \int (-X)\, \mathrm{d}(h_i \circ P_i) \right\}, \tag{11.}$$

where $\mathcal{W} = \{\tilde{w}\} \subset \mathbb{R}^m$ is a set of weights with each $\tilde{w} = (w_1, \ldots, w_m) \in \mathcal{W}$ satisfying $w_i \geq 0$ and $\sum_{i=1}^{m} w_i = 1$. The class of multiple-scenario-based distortion risk measures in Equation 11 includes the following robust subclasses:

$$\rho(X) = s \cdot \sup_{\tilde{w}\in\mathcal{W}} \left\{ \sum_{i=1}^{m} w_i \mathrm{MS}_{i,\alpha_i}(X) \right\}, \tag{12.}$$

where $\mathrm{MS}_{i,\alpha_i}(X)$ is the MS of X at confidence level α_i calculated under the ith scenario. Such a risk measure ρ is robust in both aspects: Under each scenario i, MS_{i,α_i} is statistically robust, and it incorporates multiple scenarios.

5. ELICITABILITY AND COELICITABILITY

5.1. Elicitability of One Risk Measure

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 important 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 $\mathrm{ES}(x, Y) = \int S(x, y)\mathrm{d}F_Y(y)$, where $S(x, y) : \mathbb{R}^2 \to \mathbb{R}$ is a forecasting objective function—e.g.,

[4]Gilboa & Schmeidler (1989) consider $\inf_{P\in\mathcal{P}} \int u(X)\, \mathrm{d}P$ without h_i (see also Xia 2013).

$S(x, y) = (x - y)^2$ or $S(x, y) = |x - y|$. The optimal point forecast corresponding to S is $\rho^*(F_Y) = \arg\min_x \mathrm{E}S(x, Y)$. For example, when $S(x, y) = (x - y)^2$ and $S(x, y) = |x - y|$, the optimal forecasts are the mean functional $\rho^*(F_Y) = \mathrm{E}(Y)$ 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 $\mathrm{E}S(x, Y)$. As F_Y is unknown, the expected forecasting error can be approximated by the average $\frac{1}{n}\sum_{i=1}^{n} S(x_i, Y_i)$, where $Y_1, \ldots, Y_n$ are samples 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 of the 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. The elicitability of a risk measure is also related to the statistical theory for the evaluation of probability forecasts (Lai et al. 2011).

Gneiting (2011) defines the elicitability for a set-valued statistical functional T. In the context of risk measures, we are concerned with the measurement of risk, which is a single-valued statistical functional. For example, VaR of X at level α is defined as $\mathrm{VaR}_\alpha(X) := q_\alpha^-(F_{-X})$—i.e., the left α-quantile of F_{-X}. To avoid such a minor technical nuisance, Kou & Peng (2016) slightly generalize the definition of elicitability to define the general elicitability for a single-valued statistical functional as follows.

Definition 1. 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 $S : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\rho(F) = \min\left\{x \mid x \in \arg\min_x \int S(x, y)\mathrm{d}F(y)\right\}, \ \forall F \in \mathcal{P}. \qquad 13.$$

In the definition, S is only required to satisfy the condition that $\int S(x, y)\mathrm{d}F(y)$ is well-defined and finite for any $F \in \mathcal{P}$. No other conditions, such as continuity or smoothness, are required on S. Kou & Peng (2016) show that q_α^- is general elicitable with respect to $\mathcal{D}^1 := \{F \mid F \text{ is a distribution on } \mathbb{R} \text{ and has finite first moment}\}$.

Let $\mathcal{D}_{\mathrm{disc}}$ be the set of discrete distributions having positive probabilities only on a finite number of values. Kou & Peng (2016) shows that, within the class of distortion risk measures in the form of Equation 1, VaR and the minus mean functional are the only risk measures that are general elicitable with respect to $\mathcal{D}_{\mathrm{disc}}$. In particular, VaR at level $(1 + \alpha)/2$, which is the MS at level α, provides a precise description of the average size of loss beyond VaR_α by the median of the tail loss distribution. They also show that for a fixed $\alpha_0 \in (0, 1)$, within the class of distortion risk measures, VaR_{α_0} and the minus mean functional are the only elicitable (in the sense of set-valued functionals) risk measures with respect to $\mathcal{D}_{\mathrm{disc}} \cap \{F \mid q_{\alpha_0}^-(F) = q_{\alpha_0}^+(F)\}$.

The above results imply that, among the class of risk measures in Equation 11, in order for each single-scenario risk measure $\rho_i = \int(-X)\mathrm{d}(h_i \circ P_i)$ to be elicitable, $i = 1, \ldots, m$, the only choice is

$$\rho(X) = s \cdot \sup_{\tilde{w}\in\mathcal{W}} \left\{ \sum_{i=1}^{m} w_i \mathrm{VaR}_{i,\alpha_i}(X) \right\}, \text{ and particularly, } s \cdot \sup_{\tilde{w}\in\mathcal{W}} \left\{ \sum_{i=1}^{m} w_i \mathrm{MS}_{i,\alpha_i}(X) \right\}. \qquad 14.$$

There is also literature on elicitability of convex risk measures. Weber (2006) and Gneiting (2011) show that ES is not elicitable, which "may challenge the use of the CVaR functional as a predictive measure of risk, and may provide a partial explanation for the lack of literature on the evaluation of CVaR forecasts, as opposed to quantile or VaR forecasts" (Gneiting 2011, p. 756). Weber (2006, theorem 3.1) derives a characterization theorem for risk measures with convex acceptance set $\mathcal{N}$ and convex rejection set $\mathcal{N}^c$ under two topological conditions on $\mathcal{N}$. Based on the characterization theorem of Weber (2006), Bellini & Bignozzi (2015) and Delbaen et al. (2016) provide a characterization of convex elicitable risk measures and show that the only elicitable and coherent risk measures are the expectiles. Bellini & Bignozzi (2015) make strong assumptions on the forecasting objective function $S(\cdot, \cdot)$,[5] requiring a more restrictive definition of elicitability than Gneiting (2011). The assumptions of Weber (2006) are relaxed by Delbaen et al. (2016). Liu & Wang (2021) show that the only elicitable, positively homogeneous, and monetary tail risk measures are the VaRs, and there are no elicitable tail convex or coherent risk measures except for the essential supremum.

5.2. Coelicitability of Multiple Risk Measures

The coelicitability of $k \geq 2$ statistical functionals is a weaker notion of elicitability than the notion of elicitability.

> **Definition 2.** (Lambert et al. 2008, definition 9). Single-valued statistical functionals $\rho_1(\cdot), \ldots, \rho_k(\cdot)$, $k \geq 2$, are called coelicitable 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
>
> $$(\rho_1(F), \ldots, \rho_k(F)) = \arg\min_{(x_1,\ldots,x_k)} \int S(x_1, \ldots, x_k, y)\mathrm{d}F(y), \ \forall F \in \mathcal{P}. \qquad 15.$$

The notion of coelicitability is weaker than that of elicitability because (*a*) 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 coelicitable with the corresponding function S being defined as $S(x_1, \ldots, x_k, y) := \sum_{i=1}^{k} S_i(x_i, y)$, and (*b*) if $(\rho_1, \ldots, \rho_k)$ are coelicitable, it does not imply that each ρ_i is elicitable.

Acerbi & Székely (2014) show that $(\mathrm{VaR}_\alpha(F), \mathrm{ES}_\alpha(F)) := (q_\alpha^-(F), \frac{1}{1-\alpha}\int_\alpha^1 q_s^-(F)\mathrm{d}s)$ are coelicitable with respect to a class of distributions $\mathcal{P}$ that satisfy some restrictive conditions based on an intuitive argument; Fissler & Ziegel (2016) show that $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ are coelicitable with respect to the set of loss distributions

$$\mathcal{P} = \{F \mid F \text{ has finite first moment and has unique } \alpha \text{ quantile}\},$$

and the corresponding objective function S in Definition 2 may be specified as

$$S(x_1, x_2, y) = (1_{\{x_1>y\}} - \alpha)(-G_1(-x_1) + G_1(-y)) + \frac{1}{1-\alpha} G_2(-x_2) 1_{\{x_1 \leq y\}}(y - x_1) + G_2(-x_2)(x_1 - x_2) - \mathcal{G}_2(-x_2), \qquad 16.$$

[5] These assumptions include three conditions presented by Bellini & Bignozzi (2015, definition 3.1) and two conditions in their theorem 4.4: (*a*) $S(x, y)$ is continuous in y, and (*b*) for any $x \in [-\epsilon, \epsilon]$ with $\epsilon > 0$, $S(x, y) \leq \psi(y)$ for some gauge function ψ.

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 coelicitability of (VaR_α, ES_α) implies that one can evaluate the performance of different forecasting procedures that forecast the collection of (VaR_α, ES_α) by comparing their realized forecasting errors. More precisely, procedure 1 is considered to better forecast the collection of (VaR_α, ES_α) than procedure 2 if

$$\frac{1}{T}\sum_{t=1}^{T} S(\mathrm{var}_t^1, \mathrm{es}_t^1, Y_t) < \frac{1}{T}\sum_{t=1}^{T} S(\mathrm{var}_t^2, \mathrm{es}_t^2, Y_t), \tag{17}$$

where $(\mathrm{var}_t^i, \mathrm{es}_t^i)$ 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 coelicitability of (VaR_α, ES_α) does not lead to a reliable method for evaluating forecasts for ES_α in the following sense: Even if procedure 1 better forecasts the collection (VaR_α, ES_α) than procedure 2 in the sense of Equation 17, procedure 1 may provide a much worse forecast of ES_α than procedure 2; this is illustrated in Example 5 and Example 6 in Section 6.2.

6. BACKTESTING OF A RISK MEASURE

There are three approaches for backtesting a risk measure: (*a*) The direct backtest tests if the point estimate or point forecast of the risk measurement under a model is equal to the unknown true risk measurement. (*b*) The indirect backtest can be classified into two kinds: the first kind of indirect backtest examines if the entire loss distribution, the entire tail loss distribution, or a collection of statistics including the risk measure of interest under a model is equal to the corresponding quantities under the true underlying unknown model, and the second kind of indirect backtest is based on the coelicitability of a collection of risk measures. (*c*) The forecast evaluation approach is based on the elicitability of the risk measure.

We also show the following in the subsections: (*a*) VaR and MS can be backtested by all three approaches. (*b*) There have been no direct backtesting methods for ES, and ES cannot be backtested based on elicitability. (*c*) ES cannot be reliably backtested by the two kinds of indirect backtesting methods proposed in the literature. The first kind of indirect backtesting for ES is a partial backtesting in the sense that if an indirect backtesting for ES is not rejected, it will imply that the point forecast for ES will not be rejected; however, if an indirect backtesting for ES is rejected, it will be unclear whether the point forecast for ES should be rejected. The second kind of indirect backtests, which are based on the coelicitability of (VaR_α, ES_α), cannot answer the question of whether the ES_α forecasted under a bank's model is more accurate than that forecasted under a benchmark model.

6.1. The Direct Backtesting Approach

The direct backtesting 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 1 billion. The direct backtesting approach answers the question of whether the single number 1 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 $G_{t|t-1}$ denote the bank's model of the conditional distribution of L_t given $\mathcal{F}_{t-1}$, and let $\rho^{G_{t|t-1}}(L_t)$ denote the risk measurement of L_t under the model $G_{t|t-1}$. Suppose the unknown true conditional distribution of L_t given $\mathcal{F}_{t-1}$ is $F_{t|t-1}$ and the true risk measurement is denoted as $\rho^{F_{t|t-1}}(L_t)$. Then, the direct backtesting of the risk measure ρ is to test

$$H_0 : \rho^{G_{t|t-1}}(L_t) = \rho^{F_{t|t-1}}(L_t),\ \forall t = 1, \ldots, T;\ H_1 : \text{otherwise}. \qquad 18.$$

For $\rho = \text{VaR}_\alpha = q_\alpha^-$, the null hypothesis in Equation 18 is equivalent to that $I_t := 1_{\{L_t > \text{VaR}_\alpha(L_t)\}}$, $t = 1, \ldots, T$, are i.i.d. Bernoulli$(1-\alpha)$ random variables (Christoffersen 1998, lemma 1). Based on such observations, Kupiec (1995) proposes the proportion of failure test for backtesting VaR, which is closely related to the traffic light approach of backtesting VaR adopted in the Basel Accord (Basel Comm. Bank. Superv. 1996, 2006). Christoffersen (1998) proposes conditional coverage and independence tests for VaR within a first-order Markov process model. For more recent development on the backtesting of VaR, readers are directed to Lopez (1999a,b), Engle & Manganelli (2004), Christoffersen & Pelletier (2004), Haas (2005), Campbell (2006), Christoffersen (2010), Berkowitz et al. (2011), Gaglianone et al. (2011), and Holzmann & Eulert (2014).

As $\text{MS}_\alpha = \text{VaR}_{(1+\alpha)/2}$, the backtesting of MS is exactly the same as that of VaR. In contrast, there have been no direct backtesting methods for ES in the existing literature. The reason might be simple: The null hypothesis for direct backtesting of ES is that $\text{ES}_\alpha^{G_{t|t-1}}(L_t) = \text{ES}_\alpha^{F_{t|t-1}}(L_t)$. 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 $I_t = 1_{\{L_t > \text{VaR}_\alpha(L_t)\}}$ is known under the null hypothesis for direct backtesting of VaR, and hence I_t can be used to construct a test statistic for direct backtesting of VaR.

6.2. 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 backtesting approach is to test

$$H_0 : G_{t|t-1}(x) = F_{t|t-1}(x),\ \forall x \in \mathbb{R},\ \forall t = 1, \ldots, T;\ H_1 : \text{otherwise}. \qquad 19.$$

If the null hypothesis[6] 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, this kind of indirect backtesting approach can only serve as a partial backtesting of a particular risk measure. For example, suppose a bank reports that the $\text{ES}_{99\%}$ of its trading book is 1 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 1 billion is acceptable; however, if the test is rejected, then it will be unclear if the number 1 billion should be accepted or rejected.

Strictly speaking, this indirect backtesting approach should 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 implication on the acceptability of the point forecast of a particular risk measure.

This kind of indirect backtesting approach has been proposed for backtesting ES in the literature. Berkowitz (2001) proposes likelihood ratio tests based on censored Gaussian likelihood for the test in Equation 19. Kerkhof & Melenberg (2004) propose a functional delta method for testing the hypothesis in Equation 19. Acerbi & Székely (2014) propose three indirect tests for backtesting 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:

[6]Hereafter, in all the hypothesis testing problems in this section, H_1 is taken to be not H_0.

$H_0 : G_{t|t-1,\alpha}(x) = F_{t|t-1,\alpha}(x),\ \forall x \in \mathbb{R},\ \forall t = 1, \ldots, T$, where $G_{t|t-1,\alpha}$ and $F_{t|t-1,\alpha}$ are the α-tail distribution of $G_{t|t-1}$ and $F_{t|t-1}$, respectively (see Example 2 for the definition of the α-tail distribution). The third test is the same as the test in Equation 19. All three tests proposed require that one know 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 & Curran (2015) propose an approach to indirectly backtest ES_α by testing the following: $H_0 : \int_\alpha^1 1_{\{L_t \le \mathrm{VaR}_p(L_t)\}} \mathrm{d}p, t = 1, \ldots, T$, are i.i.d., and $\mathrm{VaR}_p^{F_{t|t-1}}(L_t) = \mathrm{VaR}_p^{G_{t|t-1}}(L_t), \forall p \in [\alpha, 1), t = 1, \ldots, T$. This approach does not need to simulate random samples under the null hypothesis in order to calculate the p-value. McNeil & 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 backtest ES_α by testing the following hypothesis: $H_0 : m_t^{G_{t|t-1}} = m_t, s_t^{G_{t|t-1}} = s_t, \mathrm{VaR}_\alpha^{G_{t|t-1}}(L_t) = \mathrm{VaR}_\alpha^{F_{t|t-1}}(L_t)$, and $\mathrm{ES}_\alpha^{G_{t|t-1}}(L_t) = \mathrm{ES}_\alpha^{F_{t|t-1}}(L_t), \forall t$. This test is an indirect test for ES_α because if the null hypothesis is rejected, it is not clear if the claim $\mathrm{ES}_\alpha^{G_{t|t-1}}(L_t) = \mathrm{ES}_\alpha^{F_{t|t-1}}(L_t), \forall t$, should be rejected or not.

The second kind of indirect backtest includes those based on the coelicitability of a collection of risk measures. For example, let $(\mathrm{VaR}_\alpha^{\mathrm{Ben}}(L_t), \mathrm{ES}_\alpha^{\mathrm{Ben}}(L_t)), t = 1, \ldots, T$, be the $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ forecasted under a benchmark model such as a standard model specified by the regulator. Fissler et al. (2016) propose the two indirect backtests for ES_α based on $S(\cdot, \cdot, \cdot)$, which is the forecasting objective function defined in Equation 16 with $G_1(x) = x$ and $G_2(x) = e^x/(1 + e^x)$. The two tests are indirect backtests for ES_α because no matter whether these tests are rejected or not, we do now know whether $\mathrm{ES}_\alpha^{G_{t|t-1}}$ is more accurate than $\mathrm{ES}_\alpha^{\mathrm{Ben}}(L_t)$. 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. Assume that 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. $\mathrm{E}[S(x_1, x_2, L)]$ can be calculated analytically. Let $\mu = -1.5$, $\sigma = 1.0$, and $\alpha = 0.975$ (see Basel Comm. Bank. Superv. 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, $\mathrm{E}[S(\mathrm{VaR}_\alpha, x{\cdot}\mathrm{ES}_\alpha, L)]$, is always smaller than that of the benchmark model, $\mathrm{E}[S(x{\cdot}\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha, L)]$, 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. This is mainly due to the fact that coelicitability does not imply elicitability, and some rather strange behavior of the forecasting objective function S defined in Equation 16. This is illustrated by Kou & Peng (2016, figure 1).

Another drawback of these backtests is that the performance of the backtests further deteriorates when the scale of the loss random variable increases, because the term $G_2(-x_2)$ in Equation 16 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 backtesting ES_α. This is illustrated in Example 6.

Example 6. Suppose there is a larger bank whose loss random variable is 15 times the loss L in Example 5. Thus, the loss random variable of this larger bank has a normal distribution $\mathrm{N}(\mu, \sigma^2)$ 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, the backtests make the wrong conclusion on which model better forecasts ES_α when $x \in (0.55, 1.0)$. In addition, Kou & Peng (2016, figure 2) show 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 $\mathrm{E}[\frac{1}{1-\alpha}G_2(-x\cdot\mathrm{ES}_\alpha)1_{\{\mathrm{VaR}_\alpha<L\}}(L-\mathrm{VaR}_\alpha)+G_2(-x\cdot\mathrm{ES}_\alpha)(\mathrm{VaR}_\alpha-x\mathrm{ES}_\alpha)-\mathcal{G}_2(-x\cdot\mathrm{ES}_\alpha)]$ 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 due to some strange behavior of the forecasting objective function S defined in Equation 16.

6.3. 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 backtest 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 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 $\mathcal{D}^1 \cap \{F \mid q_\alpha^-(F) = q_\alpha^+(F)\}$, S can be defined as $S_\alpha(x,y) = (1_{\{x\geq y\}} - \alpha)(x-y)$. Then, the forecasting error is compared with a benchmark forecasting error calculated under a benchmark model to backtest VaR_α.

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

7. BASEL ACCORD RISK MEASURES

The Basel Accord risk measures are special cases of both the class of multiple-scenario-based distortion risk measures represented in Equation 11 and the class of natural risk statistics represented in Equation 9.

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 $\rho_t := s_t \max\{\frac{1}{s_t}\mathrm{VaR}_{t-1}(L_{t-1}), \frac{1}{60}\sum_{i=1}^{60}\mathrm{VaR}_{t-i}(L_{t-i})\}$, where L_{t-i} is the 10-day trading book loss starting from day $t-i$, $s_t \geq 3$ is a constant that is specified by the regulator based on the backtesting result of the institution's VaR model, and $\mathrm{VaR}_{t-i}(L_{t-i})$ is the VaR of the trading book loss at 99% confidence level calculated on day $t-i$. It is based on the information available up to and including day $t-i$ and corresponds to the calculation of VaR under the ith model, $i = 1, \ldots, 60$. Define the 61st model under which $L = 0$ with probability one. Then, the Basel II risk measure is a special case of the class of risk measures considered in Equation 12. It incorporates 61 models and two priors: One prior is $\tilde{w} = (1/s, 0, \ldots, 0, 1-1/s)$, and the other is $\tilde{w} = (1/60, 1/60, \ldots, 1/60, 0)$. It is also a special case of the natural risk statistics in Equation 9. The Basel 2.5 risk measure (Basel Comm. Bank. Superv. 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 Equations 9 and 12.

The new Basel III Accord (Basel Comm. Bank. Superv. 2019) moves from VaR to ES. The new Basel III capital charge for the market risk includes two parts, one for modelable risk factors and the other for nonmodelable risk factors. Neglecting the part for nonmodelable risk factors, the capital charge for the market risk on the tth day is specified as

$$\rho_t := s\cdot\max\left\{\frac{1}{s}\left[\rho\mathrm{ES}_{t-1}(L_{t-1}) + (1-\rho)\sum_{j=1}^{B}\mathrm{ES}_{t-1,j}(L_{t-1})\right],\right.$$
$$\left.\frac{1}{60}\sum_{i=1}^{60}\left[\rho\mathrm{ES}_{t-i}(L_{t-i}) + (1-\rho)\sum_{j=1}^{B}\mathrm{ES}_{t-i,j}(L_{t-i})\right]\right\}, \qquad 20.$$

where $s = 1.5$ or a larger number; $\rho = 0.5$; B is the number of risk factor categories, such as interest rate risk, equity risk, foreign exchange risk, commodity risk, and credit spread risk; $\mathrm{ES}_{t-i}(L_{t-i})$ is the ES at 97.5% confidence level of the loss under a stressed scenario calculated on day $t - i$; and $\mathrm{ES}_{t-i,j}(L_{t-i})$ is the ES at 97.5% confidence level of the loss under the same stressed scenario calculated on day $t - i$, with the additional hypothetical assumption that all the risk factors except the jth category of risk factors remain unchanged during the 10-day period. It is clear that the Basel III capital charge involves the calculation of ES under $60(1 + B)$ different scenarios, and it is a special case of the multiple-scenario-based distortion risk measure formulated in Equation 11 and a special case of natural risk statistics in Equation 9.

The major argument for the change from VaR_α to ES_α was that ES_α better captures tail risk than VaR_α because ES_α measures the mean size of loss beyond VaR_α, as VaR_α itself does not carry information as to the size of loss in cases when the loss does exceed VaR_α. However, ES_α is not the only risk measure that captures tail risk beyond VaR_α; in particular, MS_α also captures the size and likelihood of loss beyond VaR_α by the median of loss beyond VaR_α. The example given after Equation 3 shows that ES_α does not carry more information about the tail distribution than MS_α, simply because the mean of a distribution does not carry more information than the median of the same distribution. In addition, 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 larger than 3.44, then ES_α is smaller than MS_α (see, e.g., Von Hippel 2005).

In summary, despite lacking some mathematical convenience of subadditivity, MS_α may be preferable than ES_α as an external risk measure due to statistical and economic considerations: (*a*) MS_α also captures tail risk beyond VaR_α, (*b*) MS_α is elicitable but ES_α is not, (*c*) MS_α can be effectively backtested but ES_α cannot, (*d*) MS_α is statistically robust but ES_α is not, and (*e*) the acceptance set induced by MS_α is surplus invariant but that induced by ES_α is not.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank the anonymous reviewer for helpful comments. Xianhua Peng's research is partially supported by the General Research Fund of the Hong Kong Research Grants Council (project 16212117).

LITERATURE CITED

Acerbi C. 2002. Spectral measures of risk: a coherent representation of subjective risk aversion. *J. Bank. Finance* 26(7):1505–18

Acerbi C, Székely B. 2014. Back-testing expected shortfall. *Risk Mag.* 27(11):76–81

Adrian T, Shin HS. 2014. Procyclical leverage and value-at-risk. *Rev. Financ. Stud.* 27(2):373–403

Artzner P, Delbaen F, Eber JM, Heath D. 1999. Coherent measures of risk. *Math. Finance* 9(3):203–28

Artzner P, Delbaen F, Koch-Medina P. 2009. Risk measures and efficient use of capital. *Astin Bull.* 39(1):101–16

Basel Comm. Bank. Superv. 1996. *Amendment to the capital accord to incorporate market risks.* Work. Pap., Bank Int. Settl., Basel, Switz.

Basel Comm. Bank. Superv. 2006. *International convergence of capital measurement and capital standards: a revised framework.* Work. Pap., Bank Int. Settl., Basel, Switz.

Basel Comm. Bank. Superv. 2009. *Revisions to the Basel II market risk framework.* Work. Pap., Bank Int. Settl., Basel, Switz.

Basel Comm. Bank. Superv. 2013. *Fundamental review of the trading book: a revised market risk framework*. Work. Pap., Bank Int. Settl., Basel, Switz.

Basel Comm. Bank. Superv. 2019. *Minimal capital requirement for market risk*. Tech. Rep., Bank Int. Settl., Basel, Switz.

Bellini F, Bignozzi V. 2015. On elicitable risk measures. *Quant. Finance* 15(5):725–33

Berkowitz J. 2001. Testing density forecasts, applications to risk management. *J. Bus. Econ. Stat.* 19:465–74

Berkowitz J, Christoffersen P, Pelletier D. 2011. Evaluating value-at-risk models with desk-level data. *Manag. Sci.* 57:2213–27

Brunnermeier MK, Crockett A, Goodhart C, Persaud AD, Shin HS. 2009. *The fundamental principles of financial regulation: 11th Geneva report on the world economy*. Rep., Cent. Econ. Policy Res., London

Brunnermeier MK, Pedersen LH. 2009. Market liquidity and funding liquidity. *Rev. Financ. Stud.* 22(6):2201–38

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

Chambers CP. 2009. An axiomatization of quantiles on the domain of distribution functions. *Math. Finance* 19(2):335–42

Cheridito P, Li T. 2009. Risk measures on Orlicz hearts. *Math. Finance* 19(2):189–214

Cherny A, Madan D. 2009. New measures for performance evaluation. *Rev. Financ. Stud.* 22(7):2571–606

Christoffersen P. 1998. Evaluating interval forecasts. *Int. Econ. Rev.* 39:841–62

Christoffersen P. 2010. Backtesting. In *Encyclopedia of Quantitative Finance*, ed. R Cont. New York: Wiley. **https://doi.org/10.1002/9780470061602.eqf15018**

Christoffersen PF, Pelletier D. 2004. Backtesting value-at-risk: a duration-based approach. *J. Financ. Econom.* 2:84–108

Cont R, Deguest R, He XD. 2013. Loss-based risk measures. *Stat. Risk Model.* 30(2):133–67

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*. Work. Pap., Univ. Toronto and Bank of Montreal

Daníelsson J, Jorgensen BN, Samorodnitsky G, Sarma M, de Vries CG. 2013. Fat tails, VaR and subadditivity. *J. Econom.* 172(2):283–91

Delbaen F. 2002. Coherent risk measures on general probability spaces. In *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann*, ed. K Sandmann, PJ Schönbucher, pp. 1–37. New York: Springer

Delbaen F, Bellini F, Bignozzi V, Ziegel JF. 2016. Risk measures with the CxLS property. *Finance Stochast.* 20(2):433–53

Denneberg D. 1994. *Non-Additive Measure and Integral*. Boston: Kluwer Acad.

Dhaene J, Goovaerts MJ, Kaas R. 2003. Economic capital allocation derived from risk measures. *N. Am. Actuar. J.* 7:44–59

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

Duffie D, Pan J. 2001. Analytical value-at-risk with jumps and credit risk. *Finance Stochast.* 5(2):115–80

Eeckhoudt L, Schlesinger H. 2006. Putting risk in its proper place. *Am. Econ. Rev.* 96:280–89

Embrechts P, Wang B, Wang R. 2015. Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance Stochast.* 19:763–90

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

Engle RF, Manganelli S. 2004. CAViaR: conditional autoregressive value-at-risk by regression quantiles. *J. Bus. Econ. Stat.* 22:367–81

Fama EF, Miller MH. 1972. *The Theory of Finance*. New York: Dryden

Fissler T, Ziegel JF. 2016. Higher order elicitability and Osband's principle. *Ann. Stat.* 44(4):1680–707

Fissler T, Ziegel JF, Gneiting T. 2016. Expected shortfall is jointly elicitable with value at risk—implications for backtesting. *Risk Mag.* 29:58–61

Föllmer H, Schied A. 2002. Convex measures of risk and trading constraints. *Finance Stochast.* 6(4):429–47

Frittelli M, Gianin ER. 2002. Putting order in risk measures. *J. Bank. Finance* 26(7):1473–86

Gaglianone WP, Lima LR, Linton O, Smith DR. 2011. Evaluating value-at-risk models via quantile regression. *J. Bus. Econ. Stat.* 29(1):150–60

Garcia R, Renault É, Tsafack G. 2007. Proper conditioning for coherent VaR in portfolio management. *Manag. Sci.* 53(3):483–94

Gilboa I, Schmeidler D. 1989. Maxmin expected utility with non-unique prior. *J. Math. Econ.* 18(2):141–53

Gneiting T. 2011. Making and evaluating point forecasts. *J. Am. Stat. Assoc.* 106(494):746–62

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

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

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

Hampel FR. 1971. A general qualitative definition of robustness. *Ann. Math. Stat.* 42(6):1887–96

Hansen LP. 2013. Challenges in identifying and measuring systemic risk. In *Risk Topography: Systemic Risk and Macro Modeling*, ed. M Brunnermeier, A Krishnamurthy, pp. 15–30. Chicago: Univ. Chicago Press

Hansen LP, Sargent TJ. 2007. *Robustness*. Princeton, NJ: Princeton Univ. Press

Hart HLA. 1994. *The Concept of Law*. Oxford, UK: Clarendon. 2nd ed.

He XD, Jin H, Zhou XY. 2015. Dynamic portfolio choice when risk is measured by weighted VaR. *Math. Oper. Res.* 40(3):773–96

He XD, Peng X. 2018. Surplus-invariant, law-invariant, and positively homogeneous acceptance sets must be induced by value-at-risk. *Oper. Res.* 66(5):1268–75

Heyde CC, Kou S. 2004. On the controversy over tailweight of distributions. *Oper. Res. Lett.* 32:399–408

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

Hong CS, Herk LF. 1996. Increasing risk aversion and diversification preference. *J. Econ. Theory* 70:180–200

Huber PJ. 1981. *Robust Statistics*. New York: Wiley

Huber PJ, Ronchetti EM. 2009. *Robust Statistics*. New York: Wiley. 2nd ed.

Hull J. 2009. *Risk Management and Financial Institutions*. New York: Prentice Hall. 2nd ed.

Ibragimov R. 2004. *On the robustness of economic models to heavy-tailedness assumptions*. Tech. Rep., Yale Univ., New Haven, CT

Ibragimov R. 2009. Portfolio diversification and value at risk under thick-tailedness. *Quant. Finance* 9(5):565–80

Ibragimov R, Walden J. 2007. The limits of diversification when losses may be large. *J. Bank. Finance* 31(8):2551–69

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

Jouini E, Meddeb M, Touzi N. 2004. Vector-valued coherent risk measures. *Finance Stochast.* 8(4):531–52

Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–91

Kerkhof J, Melenberg B. 2004. Backtesting for risk-based regulatory capital. *J. Bank. Finance* 28:1845–65

Kijima M. 1997. The generalized harmonic mean and a portfolio problem with dependent assets. *Theory Decis.* 43:71–87

Koch-Medina P, Moreno-Bromberg S, Munari C. 2015. Capital adequacy tests and limited liability of financial institutions. *J. Bank. Finance* 51:93–102

Koch-Medina P, Munari C. 2016. Unexpected shortfalls of expected shortfall: extreme default profiles and regulatory arbitrage. *J. Bank. Finance* 62:141–51

Koch-Medina P, Munari C, Šikić M. 2017. Diversification, protection of liability holders and regulatory arbitrage. *Math. Financ. Econ.* 11:63–83

Kou S, Peng XH, Heyde CC. 2006. *What is a good risk measure: bridging the gaps between data, coherent risk measures, and insurance risk measures*. Work. Pap., Columbia Univ., New York

Kou SG, Peng X. 2016. On the measurement of economic tail risk. *Oper. Res.* 64(5):1056–72

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

Krätschmer V, Schied A, Zähle H. 2014. Comparative and qualitative robustness for law-invariant risk measures. *Finance Stochast.* 18(2):271–95

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

Kusuoka S. 2001. On law invariant coherent risk measures. *Adv. Math. Econ.* 3:83–95

Lai TL, Shen D, Gross S. 2011. Evaluating probability forecasts. *Ann. Stat.* 39(5):2356–82

Lambert NS, Pennock DM, Shoham Y. 2008. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce, EC '08*, pp. 129–38. New York: ACM

Liu F, Wang R. 2021. A theory for measures of tail risk. *Math. Oper. Res.* In press. **https://doi.org/10.1287/moor.2020.1072**

Lopez JA. 1999a. Methods for evaluating value-at-risk estimates. *Fed. Res. Bank San Francisco Econ. Rev.* 2:3–17

Lopez JA. 1999b. Regulatory evaluation of value-at-risk models. *J. Risk* 1:37–64

Maccheroni F, Marinacci M, Rustichini A. 2006. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica* 74(6):1447–98

McMinn RD. 1984. A general diversification theorem: a note. *J. Finance* 39:541–50

McNeil AJ, Frey R. 2000. Estimation of tail-related risk measures for heteroskedastic financial time series: an extreme value approach. *J. Empir. Finance* 7:271–300

McNeil AJ, Frey R, Embrechts P. 2005. *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton, NJ: Princeton Univ. Press

Moscadelli M. 2004. *The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee*. Work. Pap., Banca d'Italia, Rome

Osband KH. 1985. *Providing incentives for better cost forecasting*. PhD Thesis, Univ. Calif., Berkeley

Rockafellar RT, Royset J, Miranda S. 2014. Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. *Eur. J. Oper. Res.* 234(1):140–54

Rockafellar RT, Uryasev S. 2002. Conditional value-at-risk for general loss distributions. *J. Bank. Finance* 26(7):1443–71

Ruszczyński A, Shapiro A. 2006. Optimization of convex risk functions. *Math. Oper. Res.* 31(3):433–52

Samuelson PA. 1967. General proof that diversification pays. *J. Financ. Quant. Anal.* 2:1–13

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

Schmeidler D. 1986. Integral representation without additivity. *Proc. Am. Math. Soc.* 97(2):255–61

Schmeidler D. 1989. Subjective probability and expected utility without additivity. *Econometrica* 57(3):571–87

Shi Z, Werker BJM. 2012. Short-horizon regulation for long-term investors. *J. Bank. Finance* 36(12):3227–38

So MKP, Wong CM. 2012. Estimation of multiple period expected shortfall and median shortfall for risk management. *Quant. Finance* 12(5):739–54

Song Y, Yan JA. 2006. The representations of two types of functionals on $l^\infty(\omega, \mathcal{F})$ and $l^\infty(\omega, \mathcal{F}, \mathbb{P})$. *Sci. China Ser. A* 49:1376–82

Song Y, Yan JA. 2009. Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. *Insurance Math. Econ.* 45(3):459–65

Staum J. 2013. Excess invariance and shortfall risk measures. *Oper. Res. Lett.* 41(1):47–53

Tasche D. 2002. Expected shortfall and beyond. *J. Bank. Finance* 26(7):1519–33

Thomson W. 1979. Eliciting production possibilities from a well-informed manager. *J. Econ. Theory* 20(3):360–80

Tversky A, Kahneman D. 1992. Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertainty* 5:297–323

Von Hippel PT. 2005. Mean, median, and skew: correcting a textbook rule. *J. Stat. Educ.* 13(2):1–13

Wang R, Zitikis R. 2021. An axiomatic foundation for the expected shortfall. *Manag. Sci.* 67(3):1413–29

Wang SS, Young VR, Panjer HH. 1997. Axiomatic characterization of insurance prices. *Insurance Math. Econ.* 21(2):173–83

Weber S. 2006. Distribution-invariant risk measures, information, and dynamic consistency. *Math. Finance* 16(2):419–42

Weber S, Anderson W, Hamm AM, Knispel T, Liese M, Salfeld T. 2013. Liquidity-adjusted risk measures. *Math. Financ. Econ.* 7(1):69–91

Wen Z, Peng X, Liu X, Bai X, Sun X. 2013. *Asset allocation under the Basel Accord risk measures*. Work. Pap., Peking Univ., Beijing, China

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

Xia J. 2013. *Comonotonic convex preferences.* Work. Pap., Acad. Math. Systems Sci., Chinese Acad. Sci., Beijing

Yaari ME. 1987. The dual theory of choice under risk. *Econometrica* 55:95–115