



Research Paper

What is the best risk measure in practice? A comparison of standard measures

Susanne Emmer,¹ Marie Kratz¹ and Dirk Tasche²

¹ESSEC Business School, CREAR, Avenue Bernard Hirsch, 95021 Cergy-Pontoise Cedex, France; emails: susanne.emmer@yahoo.de, kratz@essec.edu

²Prudential Regulation Authority, Bank of England, 20 Moorgate, London EC2R 6DA, UK; email: dirk.tasche@gmx.net

(Received January 27, 2014; revised February 1, 2015; accepted February 5, 2015)

ABSTRACT

Expected shortfall (ES) has been widely accepted as a risk measure that is conceptually superior to value-at-risk (VaR). At the same time, however, it has been criticized for issues relating to backtesting. In particular, ES has been found not to be elicitable, which means that backtesting for ES is less straightforward than, for example, backtesting for VaR. Expectiles have been suggested as potentially better alternatives to both ES and VaR. In this paper, we revisit the commonly accepted desirable properties of risk measures such as coherence, comonotonic additivity, robustness and elicibility. We check VaR, ES and expectiles with regard to whether or not they enjoy these properties, with particular emphasis on expectiles. We also consider their impact on capital allocation, an important issue in risk management. We find that, despite the caveats that apply to the estimation and backtesting of ES, it can be considered a good risk measure. As a consequence, there is no sufficient evidence to justify an all-inclusive replacement of ES by expectiles in applications. For backtesting ES, we propose an empirical approach that consists of replacing ES by a set of four quantiles, which should allow us to make use of backtesting methods for VaR.

Keywords: backtesting; capital allocation; coherence; diversification; conditional elicibility; risk measures.

1 INTRODUCTION

Risk management is a core competence of financial institutions such as banks, insurance companies and investment funds, among others. Techniques for the measurement of risk are clearly central to the process of managing risk. Risk can be measured in terms of probability distributions. However, it is sometimes useful to express risk with one number that can be interpreted as a capital amount. Tools that map loss distributions or random variables to capital amounts are called risk measures. The following questions are of crucial importance for financial institutions.

- What properties should we expect from a risk measure?
- What is a “good” risk measure?
- Does a “best” risk measure exist?

Much research in economics, finance and mathematics has been devoted to answering these questions. Cramér was one of the earliest researchers on risk capital, introducing ruin theory (Cramér 1930). A major contribution was also made by Markowitz (1952) with modern portfolio theory. The variance of the profit-and-loss (P&L) distribution then became the dominating risk measure in finance. But using this risk measure has two important drawbacks. It requires that the risks are random variables with finite variance. It also implicitly assumes that their distributions are approximately symmetric around the mean, since the variance does not distinguish between positive and negative deviations from the mean. Since then, many risk measures have been proposed, of which value-at-risk (VaR) and expected shortfall (ES) seem to be the most popular.

In the seminal work by Artzner *et al* (1999), desirable properties of risk measures were formalized in a set of axioms. Because ES has the important property of coherence, it has replaced VaR, which does not satisfy this property in all cases, in many institutions for risk management and, in particular, for capital allocation (Tasche 2008). The Basel Committee on Banking Supervision also recommends replacing VaR with ES in internal market risk models (Basel Committee on Banking Supervision 2013). Recently, a study by Gneiting (2011) has pointed out that there could be an issue with the direct backtesting of ES estimates because ES is not “elicitable”. Therefore, with regard to the feasibility of backtesting, recent studies (Bellini *et al* 2014; Ziegel 2014) have suggested expectiles as coherent and elicitable alternatives to ES (see also Chen (2014) for a detailed discussion of the issue).

The aim of this paper is to provide a compendium of some popular risk measures based on probability distributions, in order to discuss and compare their properties. We can then provide answers to the questions raised above and study the impact of the choice of risk measure in terms of risk management and model validation. Several

recent review papers (see, for example, Embrechts and Hofert 2014a; Embrechts *et al* 2014) also discuss the use of risk measures and some of their properties in the context of regulation. Here, we present a panorama of the mathematical properties of four standard risk measures, variance, VaR, ES and expectile, addressed to both academics and professionals.

We consider a portfolio of m risky positions, where $L_i, i \in \{1, \dots, m\}$, represents the loss in the i th position. Then, in the generic one-period loss model, the portfolio-wide loss is given by $L = \sum_{i=1}^m L_i$. In this model, losses are positive numbers, whereas gains are negative numbers. We assume that the portfolio loss variable L is defined on a probability space (Ω, \mathcal{F}, P) .

The paper is organized as follows. After this introductory section, Section 2 recalls the main definitions and properties of what is expected from a risk measure, such as coherence, comonotonic additivity, law invariance, elicibility and robustness, before presenting the three downside risk measures that we want to evaluate in this study. In Section 3, we compare these risk measures with respect to their properties, starting with an overview. After summing up the most important results about the subadditivity of VaR, we look at different concepts of robustness, discuss the elicibility of ES and expectiles, and observe that expectiles are not comonotonically additive. Section 4 deals with capital allocation and diversification benefits, which are important areas of application for risk measures and for risk management. We recall the definition of risk contributions of risky positions to portfolio-wide risk and show how to compute risk contributions for expectiles. Further, we introduce the concept of a diversification index for the quantification and comparison of the diversification of portfolios. In Section 5, we present methods for backtesting in general and look in more detail at ES. The paper ends with Section 6, which provides a discussion of the advantages and disadvantages of the different risk measures and a recommendation for the choice of a risk measure in practice.

Notation

$\mathbf{1}_M$ denotes the indicator function of the set M , ie, $\mathbf{1}_M(x) = 1$ if $x \in M$, and $\mathbf{1}_M(x) = 0$ if $x \notin M$.

2 RISK MEASURES: DEFINITION AND BASIC PROPERTIES

Risk and risk measure are terms that have no unique definition or usage. It would be natural to measure risk in terms of probability distributions. However, often it is useful to express risk with one number. Mappings from spaces of probability distributions or random variables into real numbers are called risk measures. In this paper, a risk

measure is understood as providing a risk assessment in the form of a capital amount that serves as some kind of buffer against unexpected future losses.¹

2.1 Coherence and related properties

Artzner *et al* (1999) demonstrate that, given some “reference instrument”, there is a natural way to define a measure of risk by describing how close or far a position is from acceptance by the regulator. In the context of Artzner *et al* (1999), the set of all risks is the set of all real-valued functions on a probability space Ω , which is assumed to be finite. Artzner *et al* (1999) define “the measure of risk of an unacceptable position *once a reference, prudent, investment has been specified* as the minimum extra capital . . . which, invested in the reference instrument, makes the future value of the modified position become acceptable”. Artzner *et al* (1999) call the investor’s future net worth “risk”. Moreover, they state four axioms that any risk measure used for effective risk regulation and management should satisfy. Such risk measures are then said to be coherent. Coherence bundles certain mathematical properties that are possible criteria for the choice of a risk measure.

DEFINITION 2.1 A risk measure ρ is called “coherent” if it satisfies the following conditions.

- Homogeneity: ρ is homogeneous if for all loss variables L and $h \geq 0$ it holds that

$$\rho(hL) = h\rho(L).$$

- Subadditivity: ρ is subadditive if for all loss variables L_1 and L_2 it holds that

$$\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2).$$

- Monotonicity: ρ is monotonic if for all loss variables L_1 and L_2 it holds that

$$L_1 \leq L_2 \Rightarrow \rho(L_1) \leq \rho(L_2).$$

- Translation invariance: ρ is translation invariant if for all loss variables L and $a \in \mathbb{R}$ it holds that

$$\rho(L - a) = \rho(L) - a.$$

Comonotonic additivity is another property of risk measures that is mainly of interest as a complementary property to subadditivity.

¹ See Rockafellar and Uryasev (2013) for alternative interpretations of risk measures.

DEFINITION 2.2 Two real-valued random variables L_1 and L_2 are said to be comonotonic if there exists a real-valued random variable X (the common risk factor) and nondecreasing functions f_1 and f_2 such that

$$L_1 = f_1(X) \quad \text{and} \quad L_2 = f_2(X).$$

A risk measure ρ is comonotonically additive if for any comonotonic random variables L_1 and L_2 it holds that

$$\rho(L_1 + L_2) = \rho(L_1) + \rho(L_2).$$

Comonotonicity may be considered the strongest possible dependence of random variables (Embrechts *et al* 2002). Hence, if a risk measure is both subadditive and comonotonically additive then, on the one hand, it rewards diversification (via subadditivity) but, on the other hand, it does not attribute any diversification benefits to comonotonic risks (via comonotonic additivity). This seems quite intuitive. Risk measures that depend only on the distribution of losses are of special interest, because their values can be estimated from loss observations only (ie, no additional information such as stress scenarios is needed).

DEFINITION 2.3 A risk measure ρ is law invariant if

$$P(L_1 \leq \ell) = P(L_2 \leq \ell), \quad \ell \in \mathbb{R} \Rightarrow \rho(L_1) = \rho(L_2).$$

2.2 Elicitability

An interesting criterion when estimating a risk measure is elicibility, which was introduced by Osband (1985) and Lambert *et al* (2008), and then Gneiting (2011). We briefly recall its definition, which is linked to that of scoring functions. For further details we refer the reader to the recent review on probabilistic forecasting, including the notion of elicibility, by Gneiting and Katzfuss (2014). For the definition of elicibility, we first introduce the concept of strictly consistent scoring functions.

A scoring function aims at assigning a numerical score to a single-valued point forecast based on the predictive point and realization.

DEFINITION 2.4 A scoring function is a function

$$\begin{aligned} s: \mathbb{R} \times \mathbb{R} &\rightarrow [0, \infty), \\ (x, y) &\rightarrow s(x, y), \end{aligned}$$

where x and y are the point forecasts and observations, respectively.

DEFINITION 2.5 Let ν be a functional on a class of probability measures \mathcal{P} on \mathbb{R} :

$$\begin{aligned} \nu: \mathcal{P} &\rightarrow 2^{\mathbb{R}} \quad (\text{the power set of } \mathbb{R}), \\ P &\mapsto \nu(P) \subset \mathbb{R}. \end{aligned}$$

A scoring function $s: \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty)$ is consistent for the functional ν relative to the class \mathcal{P} if and only if, for all $P \in \mathcal{P}$, $t \in \nu(P)$ and $x \in \mathbb{R}$,

$$\mathbb{E}_P[s(t, L)] \leq \mathbb{E}_P[s(x, L)],$$

where L is a real-valued random variable with distribution P .

The function s is strictly consistent if it is consistent and

$$\mathbb{E}_P[s(t, L)] = \mathbb{E}_P[s(x, L)] \Rightarrow x \in \nu(P).$$

DEFINITION 2.6 The functional ν is elicitable relative to \mathcal{P} if and only if there is a scoring function s that is strictly consistent for ν relative to \mathcal{P} .

EXAMPLE 2.7 Standard examples of scoring functions are the following:

- (a) $s(x, y) = (x - y)^2$,
- (b) $s(x, y) = (\mathbf{1}_{\{x \geq y\}} - \tau)(x - y)^2 \operatorname{sgn}(x - y)$, $0 < \tau < 1$ fixed,
- (c) $s(x, y) = |x - y|$,
- (d) $s(x, y) = (\mathbf{1}_{\{x \geq y\}} - \alpha)(x - y)$, $0 < \alpha < 1$ fixed.

The (a) squared, (b) weighted squared, (c) absolute and (d) weighted absolute errors above are strictly consistent scoring functions; the mean functional is elicited by the squared error, the expectile is elicited by the weighted squared error, the median is elicited by the absolute error and the quantile is elicited by the weighted absolute error (see Newey and Powell 1987).

Elicitability is a helpful criterion for the determination of optimal point forecasts; the class of (strictly) consistent scoring functions for a functional is identical to the class of functions under which (only) the functional is an optimal point forecast. Hence, if we have found a strictly consistent scoring function for a functional ν , we can determine the optimal forecast \hat{x} for $\nu(P)$ by

$$\hat{x} = \arg \min_x \mathbb{E}_P[s(x, L)].$$

Hence, the elicibility of a functional of probability distributions may be interpreted as the property that the functional can be estimated by generalized regression. Another property that makes elicibility an important concept is that it can be used to compare the performance of different forecast methods (see Gneiting (2011) for a detailed discussion).

2.3 Conditional elicibility

So far, we have only distinguished between elicitable and nonelicitable functionals. However, it turns out that some useful risk measures are not elicitable but rather “second-order” elicitable in the following sense.

DEFINITION 2.8 (Conditional elicibility) A functional ν of \mathcal{P} is called “conditionally elicitable” if there exist functionals $\tilde{\gamma}$ and $\gamma: \mathcal{D} \rightarrow 2^{\mathbb{R}}$ with $\mathcal{D} \subset \mathcal{P} \times 2^{\mathbb{R}}$ such that

- (i) $\tilde{\gamma}$ is elicitable relative to \mathcal{P} ,
- (ii) $(P, \tilde{\gamma}(P)) \in \mathcal{D}$ for all $P \in \mathcal{P}$,
- (iii) for all $c \in \tilde{\gamma}(\mathcal{P})$, the functional $\gamma_c: \mathcal{P}_c \rightarrow 2^{\mathbb{R}}$, $P \mapsto \gamma(P, c) \subset \mathbb{R}$ is elicitable relative to $\mathcal{P}_c = \{P \in \mathcal{P} : (P, c) \in \mathcal{D}\}$,
- (iv) $\nu(P) = \gamma(P, \tilde{\gamma}(P))$ for all $P \in \mathcal{P}$.

Sometimes, c and $\gamma(P, c)$, respectively, are single valued. In this case, we identify the one-point sets c and $\gamma(P, c)$, respectively, with their unique elements.

Conditional elicibility is a helpful concept for the forecasting of some risk measures that are not elicitable. In Section 3.3, we will study ES as an example of a risk measure whose conditional elicibility provides us with the possibility to forecast it in two steps. Indeed, due to the elicibility of $\tilde{\gamma}$, we can first forecast $\tilde{\gamma}(P)$ and then, in a second step, take the result for $\tilde{\gamma}(P)$ as fixed and forecast $\gamma(P, c)$, due to the elicibility of γ_c .

With regard to backtesting and forecast comparison, conditional elicibility offers a way of splitting up a forecast method into two component methods and separately backtesting and comparing their forecast performances. This reflects an approach often applied in practice where a complex forecast method is decomposed into component methods that are separately validated. While this approach is attractive for making complex issues tractable, it does not necessarily entail the optimal choice of forecast models.

REMARK 2.9 Every elicitable functional is conditionally elicitable.

2.4 Robustness

Another important issue when estimating risk measures is robustness. Without robustness (defined in an appropriate sense), results may not be meaningful, since small measurement errors in the loss distribution can have a huge impact on the estimate of the risk measure. This is why we investigate robustness in terms of continuity. Since

most of the relevant risk measures are not continuous with respect to the weak topology, we need a stronger notion of convergence. Therefore, and due to some scaling properties that are convenient in risk management, it is useful to consider the Wasserstein distance when investigating the robustness of risk measures (see, for example, Bellini *et al* 2014).

Recall that the Wasserstein distance between two probability measures P and Q is defined as follows:

$$d_W(P, Q) = \inf\{E(|X - Y|) : X \sim P, Y \sim Q\}.$$

When we call a risk measure robust with respect to the Wasserstein distance, we mean continuity with respect to the Wasserstein distance in the following sense.

DEFINITION 2.10 Let P_n , $n \geq 1$, and P be probability measures, and let $X_n \sim P_n$, $n \geq 1$, and $P \sim X$. A risk measure ρ is called continuous at X with respect to the Wasserstein distance if

$$\lim_{n \rightarrow \infty} d_W(X_n, X) = 0 \Rightarrow \lim_{n \rightarrow \infty} |\rho(X_n) - \rho(X)| = 0.$$

In Section 3.2, we discuss the robustness properties of some popular risk measures with regard to the Wasserstein distance.

Cont *et al* (2010) use a different, potentially more intuitive, concept of robustness that takes the estimation procedure into account. They investigate robustness as the sensitivity of the risk measure estimate to the addition of a new data point to the data set, which is used as the basis for estimation. It turns out that for the same risk measure the estimation method can have a significant impact on sensitivity. For example, the risk measure estimate can react in a completely different way on an additional data point if we fit a parametric model instead of using the empirical loss distribution. Thus, robustness in the sense of Cont *et al* (2010) relates more to sensitivity to outliers in the data sample than to mere measurement errors. Cont *et al* (2010) also show that there is a conflict between the subadditivity and robustness of a risk measure.

In contrast to robustness based on continuity with respect to weak topology or Wasserstein distance, the concept of Cont *et al* (2010) allows us to distinguish between different degrees of robustness. This concept may make it hard to decide whether or not a risk measure is still reasonably risk sensitive or no longer robust with respect to data outliers in the estimation sample. However, in finance and insurance, large values do occur and are not outliers or measurement errors, but rather facts that are a part of the observed process itself. In particular, in (re)insurance, one could argue that large claims are actually more accurately monitored than small ones, and their values better estimated. Thus, the question of robustness in the sense of Cont *et al* (2010) may not be so relevant in this context. That is why, for the purpose of this paper, we adopt a notion of robustness based on the Wasserstein distance, which focuses on small measurement errors.

2.5 Popular risk measures

Variance and standard deviation were historically the dominating risk measures in finance. However, in the past twenty years or so, they have often been replaced in practical applications by VaR, which is currently the most popular downside risk measure.

DEFINITION 2.11 The VaR at level $\alpha \in (0, 1)$ of a loss variable L is defined as the α -quantile of the loss distribution

$$\text{VaR}_\alpha(L) = q_\alpha(L) = \inf\{\ell: P(L \leq \ell) \geq \alpha\}.$$

VaR is sometimes criticized for a number of different reasons. Most important are its lack of the subadditivity property and the fact that it completely ignores the severity of losses in the far tail of the loss distribution. The coherent risk measure ES was introduced to solve these issues.

DEFINITION 2.12 (Acerbi and Tasche 2002) The ES at level $\alpha \in (0, 1)$ (also called tail VaR or superquantile) of a loss variable L is defined as

$$\begin{aligned} \text{ES}_\alpha(L) &= \frac{1}{1-\alpha} \int_\alpha^1 q_u(L) \, du \\ &= \mathbb{E}[L \mid L \geq q_\alpha(L)] \\ &\quad + (\mathbb{E}[L \mid L \geq q_\alpha(L)] - q_\alpha(L)) \left(\frac{P[L \geq q_\alpha(L)]}{1-\alpha} - 1 \right). \end{aligned} \quad (2.1)$$

If $P[L = q_\alpha(L)] = 0$ (in particular, if the distribution of L is continuous), $\text{ES}_\alpha(L) = \mathbb{E}[L \mid L \geq q_\alpha(L)]$.

2.6 Expectiles

ES has been shown not to be elicitable (Gneiting 2011). That is why expectiles have been suggested as coherent and elicitable alternatives (Bellini *et al* 2014; Ziegel 2014). The following definition characterizes expectiles analogously to the familiar characterization of expected values as solutions to minimization problems. As such, they generalize expected values. However, this definition is not the most general, because it requires the random variable to be square integrable. Therefore, we revise it afterwards.

DEFINITION 2.13 For $0 < \tau < 1$ and square integrable L , the τ -expectile $e_\tau(L)$ is defined as

$$e_\tau(L) = \arg \min_{\ell \in \mathbb{R}} \mathbb{E}[\tau \max(L - \ell, 0)^2 + (1 - \tau) \max(\ell - L, 0)^2].$$

Since VaR is not coherent and ES lacks direct elicibility, it is interesting to look for risk measures that are coherent as well as elicitable. A possible candidate is the expectile, which we just defined; a more general but less intuitive definition is suggested by the following observation.

LEMMA 2.14 (Newey and Powell 1987; Bellini *et al* 2014) *If L is an integrable random variable, then $e_\tau(L)$ is the unique solution ℓ of*

$$\tau \mathbb{E}[\max(L - \ell, 0)] = (1 - \tau) \mathbb{E}[\max(\ell - L, 0)].$$

Consequently, $e_\tau(L)$ satisfies

$$e_\tau(L) = \frac{\tau \mathbb{E}[L \mathbf{1}_{\{L \geq e_\tau(L)\}}] + (1 - \tau) \mathbb{E}[L \mathbf{1}_{\{L < e_\tau(L)\}}]}{\tau P[L \geq e_\tau(L)] + (1 - \tau) P[L < e_\tau(L)]}.$$

According to Gneiting (2011, Theorem 10), expectiles are elicitable on the space of all integrable random variables.

PROPOSITION 2.15 (Bellini *et al* 2014) *Expectiles have the following properties.*

- (i) *For $0 < \tau < 1$, expectiles are homogeneous and law invariant. As a consequence, expectiles are additive for linearly dependent random variables, ie,*

$$\text{corr}[L_1, L_2] = 1 \Rightarrow e_\tau(L_1 + L_2) = e_\tau(L_1) + e_\tau(L_2).$$

- (ii) *For $\frac{1}{2} \leq \tau < 1$, expectiles are subadditive (and hence coherent), whereas, for $\frac{1}{2} \geq \tau > 0$, they are superadditive.*

Ziegel (2014) has recently shown that expectiles are indeed the only law-invariant and coherent elicitable risk measures.

From Lemma 2.14 and Proposition 2.15, it seems as if expectiles are ideal for making good the deficiencies of VaR and ES. This is not the case, however, because expectiles are not comonotonically additive, as follows immediately from their so-called Kusuoka representation, which is given, for example, in Ziegel (2014).

PROPOSITION 2.16 *For $\frac{1}{2} < \tau < 1$, expectiles are not comonotonically additive.*

PROOF OF PROPOSITION 2.16 *If e_τ were comonotonically additive, then, by Tasche (2002, Theorem 3.6), it would be a so-called spectral risk measure. But then, by Ziegel (2014, Corollary 4.3), it would not be elicitable, in contradiction to Gneiting (2011, Theorem 10). \square*

TABLE 1 Properties of standard risk measures.

Property	Variance	VaR	ES	e_τ (for $\tau \geq \frac{1}{2}$)
Coherence			×	×
Comonotonic additivity		×	×	
Robustness with respect to weak topology		×		
Robustness with respect to Wasserstein distance	×	×	×	×
Elicitability		×		×
Conditional elicibility	×	×	×	×

*It can be shown that VaR at level α is robust with respect to the weak topology at F_0 if F_0^{-1} is continuous at α . See, for example, Huber and Ronchetti (2009, Theorem 3.7).

3 PROPERTIES OF THE STANDARD RISK MEASURES

Although considering different risk measures would give a more complete picture of the riskiness of a portfolio, in practice one often has to choose just one number, which should be reported as the basis for strategic decisions. To help with this choice, let us start with Table 1, which provides an overview of the considered risk measures and their properties. We come back to these in more detail later.

3.1 When is value-at-risk subadditive?

The subadditivity property fails to hold for VaR in general, so VaR is not a coherent measure. The lack of subadditivity contradicts the notion that there should be a diversification benefit associated with merging portfolios. As a consequence, a decentralization of risk management using VaR is difficult, since we cannot be sure that by aggregating VaR numbers for different portfolios or business units we will obtain a bound for the overall risk of the enterprise. Moreover, in contrast to ES at the same confidence level, VaR at level α gives no information about the severity of tail losses, which occur with a probability of less than $1 - \alpha$.

When looking at aggregated risks $\sum_{i=1}^n L_i$, it is well known (Acerbi and Tasche 2002) that the risk measure ES is coherent. In particular, it is subadditive. In contrast, VaR is not subadditive in general. Indeed, examples (see, for example, Embrechts *et al* 2009a) can be given where it is superadditive, ie,

$$\text{VaR}_\alpha\left(\sum_{i=1}^n L_i\right) > \sum_{i=1}^n \text{VaR}_\alpha(L_i).$$

Whether or not VaR is subadditive depends on the properties of the joint loss distribution.

We will not provide an exhaustive review of results on conditions for the subadditivity of VaR. We present only three of these results in the remainder of this section, namely three standard cases.

- (i) The random variables are independent and identically distributed (iid) as well as positively regularly varying.
- (ii) The random variables have an elliptical distribution.
- (iii) The random variables have an Archimedean survival dependence structure.

For further related results, see, for example, Daniélsion *et al* (2013) and Embrechts *et al* (2002, 2009a,b, 2013).

Ad (i). The following result presents a condition on the tail behavior of iid random variables for VaR to satisfy asymptotic subadditivity.

PROPOSITION 3.1 (Embrechts *et al* 2009a) *Consider iid random variables $X_i, i = 1, \dots, n$, with common cumulative distribution function F . Assume they are regularly varying with tail index $\beta > 0$, which means that the right tail $1 - F$ of their distribution satisfies*

$$\lim_{x \rightarrow \infty} \frac{1 - F(ax)}{1 - F(x)} = a^{-\beta} \quad \text{for all } a > 0.$$

Then the risk measure VaR is asymptotically subadditive for X_1, \dots, X_n if and only if $\beta \geq 1$:

$$\lim_{\alpha \nearrow 1} \frac{\text{VaR}_\alpha(\sum_{i=1}^n X_i)}{\sum_{i=1}^n \text{VaR}_\alpha(X_i)} \leq 1 \Leftrightarrow \beta \geq 1.$$

Ad (ii). Another important class of distributions that implies the subadditivity of VaR is the class of elliptical distributions.

PROPOSITION 3.2 (Embrechts *et al* 2002) *Let $X = (X_1, \dots, X_n)$ be a random vector having an elliptical distribution. Consider the set of linear portfolios*

$$M = \left\{ Z = \sum_{i=1}^n \lambda_i X_i \mid \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Then, VaR at level α is subadditive on M if $0.5 < \alpha < 1$:

$$\text{VaR}_\alpha(Z_1 + Z_2) \leq \text{VaR}_\alpha(Z_1) + \text{VaR}_\alpha(Z_2), \quad Z_1, Z_2 \in M.$$

Ad (iii). Further, there exists an analogous result for another type of dependence, the Archimedean survival copula.

PROPOSITION 3.3 (Embrechts *et al* 2009b) *Consider random variables X_i , $i = 1, \dots, n$, which have the same continuous marginal distribution function F . Assume the tail distribution $\bar{F} = 1 - F$ is regularly varying with tail index $-\beta < 0$, ie, $\bar{F}(x) = x^{-\beta} G(x)$ for some function G slowly varying at infinity, and assume $(-X_1, \dots, -X_n)$ has an Archimedean copula with generator Ψ , which is regularly varying at 0 with index $-\alpha < 0$. Then, for all $\alpha > 0$, we have*

- *VaR is asymptotically subadditive for all $\beta > 1$,*
- *VaR is asymptotically superadditive for all $\beta < 1$.*

Recently, numerical and analytical techniques have been developed in order to evaluate the risk measures VaR and ES under different dependence assumptions regarding the loss random variables. Such techniques certainly help for a better understanding of the aggregation and diversification properties of risk measures, in particular of noncoherent measures such as VaR. In this paper, we do not review all these techniques and results but refer to Embrechts *et al* (2013) and the references therein for an overview.

Nevertheless, it is worth mentioning two recent studies: a new numerical algorithm introduced by Embrechts *et al* (2013) to provide bounds for the VaR of aggregated risks, and a study by Kratz (2013, 2014) on the evaluation of the VaR of aggregated heavy-tailed risks. The numerical algorithm introduced in Embrechts *et al* (2013) allows for the computation of reliable lower and upper bounds for the VaR of high-dimensional (inhomogeneous) portfolios, whatever the dependence structure. To quote the authors:

...additional positive dependence information will typically not improve the upper bound substantially. In contrast higher order marginal information on the model, when available, may lead to strongly improved bounds.

This is a good news since, in practice, typically only the marginal loss distribution functions are known or statistically estimated, while the dependence structure between the losses is either completely or partially unknown. In Kratz (2014), a new approach, called Normex, is developed to provide accurate estimates of high quantiles for aggregated independent heavy-tailed risks. This method depends only weakly upon the sample size and gives good results for any nonnegative tail index of the risks.

3.2 Robustness

With respect to the weak topology, most of the common risk measures are discontinuous. Therefore, and due to some convenient scaling properties detailed in Stahl *et al* (2012, Proposition 2.1), in risk management one usually considers robustness

as continuity with respect to the Wasserstein distance, as defined by (2.4). According to Stahl *et al* (2012), variance, ES, expectiles and mean are discontinuous with respect to the weak topology, whereas VaR at the level α is robust at F_0 if F_0^{-1} is continuous at α . Stahl *et al* (2012) observe that mean, VaR and ES are continuous with respect to the Wasserstein distance, and Bellini *et al* (2014) show that expectiles are Lipschitz-continuous with respect to the Wasserstein distance with constant

$$K = \max \left\{ \frac{\alpha}{1-\alpha}; \frac{1-\alpha}{\alpha} \right\},$$

which implies continuity with respect to the Wasserstein distance.

With regard to robustness in the sense given in Cont *et al* (2010) (as mentioned in Section 2.4), Cont *et al* demonstrate that historical ES is much more sensitive to the addition of a data point than VaR. Moreover, in contrast to VaR, ES is sensitive to the data point's size. The authors also investigate the impact of the estimation method on the sensitivity and find that historical ES at the 99% level is much more sensitive than Gaussian and Laplace ES. Moreover, they discuss a potential conflict between the requirements of subadditivity, and therefore also coherence, and robustness of a risk measure estimate.

Taking into account that VaR, because of its definition as a quantile, is insensitive to the sizes of data points that do not fall into a neighborhood of VaR, the observations by Cont *et al* (2010) are not too surprising. The notion of ES was introduced precisely as a remedy to the lack of risk sensitivity of VaR.²

Finally, note that, in practice, the estimation of ES will often be based on larger subsamples than the estimation of VaR. For instance, when using 100 000 simulation iterations, ES at the 99% level is estimated with 1000 points, while the VaR estimate is based on a small neighborhood of the 99 000th order statistic. Moreover, when investigating empirically the scaling properties of the VaR and ES of aggregated financial returns, Hauksson *et al* (2001) noted that the numerical stability of the scaling exponent was much higher with ES. This observation, in a way, counters the comments of Cont *et al* (2010) with regard to the amount of data needed for estimation. For, often, one can use high-frequency data to precisely estimate ES and then use the scaling property to determine ES for aggregated risks.

3.3 Elicibility and conditional elicibility

The lack of coherence of VaR, which has long been the most popular risk measure in practice, draws attention to another downside risk measure, ES, as defined in (2.1).

² Recently, Jadhav *et al* (2013) suggested that “modified ES” was a robust and coherent variation of ES. However, their proof of coherence is wrong. Moreover, Jadhav *et al* seem to have overlooked that Cont *et al* (2010, Section 3.2.3) looked at modified ES before and observed that it is not coherent.

ES is a coherent risk measure and, in contrast to VaR, it is sensitive to the severity of losses beyond VaR. Nevertheless, when it comes to forecasting and backtesting ES, a potential deficiency arises that is not present in VaR. Gneiting (2011) showed that ES is not elicitable. He proved that the existence of convex level sets is a necessary condition for the elicibility of a risk measure and disproved the existence of convex level sets for ES. It is interesting to note that other important risk measures, such as the variance, are not elicitable either (Lambert *et al* 2008).

LEMMA 3.4 *For continuous distributions with finite means, ES is conditionally elicitable.*

PROOF OF LEMMA 3.4 Fix $\alpha \in (0, 1)$. Let $\mathcal{P} = \{\text{continuous distributions on } \mathbb{R} \text{ with finite means}\}$ and

$$\mathcal{D} = \{(P, c) \in \mathcal{P} \times \mathbb{R} : P([c, \infty)) > 0\}.$$

For continuous distributions P , ES simplifies to $ES_\alpha(L) = \mathbb{E}[L \mid L \geq q_\alpha(L)]$, where L denotes a generic random variable with distribution P . Hence, we can rewrite $ES_\alpha(L)$ using $\gamma : \mathcal{D} \rightarrow \mathbb{R}$, defined by

$$(P, c) \mapsto \gamma(P, c) := E_P[L \mid L \geq c],$$

and $\tilde{\gamma} : \mathcal{P} \rightarrow \mathbb{R}$, defined by

$$P \mapsto \tilde{\gamma}(P) := q_\alpha(L).$$

Since we have $P(L \geq q_\alpha(L)) = 1 - \alpha > 0$ for continuous distributions P , properties (ii) and (iv) of Definition 2.8 are satisfied. Property (i) holds because quantiles of distributions with finite means are elicitable with strictly consistent scoring function $s(x, y) = (\mathbf{1}_{\{x \geq y\}} - \alpha)(x - y)$ (Newey and Powell 1987). For fixed $c \in \mathbb{R}$ and \mathcal{P}_c defined as in Definition 2.8, an application of Gneiting (2011, Theorem 7) shows that γ_c is elicitable with strictly consistent scoring function

$$s(x, y) = (\phi(y) - \phi(x) - \phi'(x)(y - x))\mathbf{1}_{[c, \infty)}(x), \quad \text{where } \phi(x) = \frac{x^2}{1 + |x|}.$$

This proves property (iii) of Definition 2.8. □

In practice, Lemma 3.4 implies that, due to its conditional elicibility, we can try to forecast ES in a two-step procedure.

- (1) We forecast the quantile as

$$\hat{q}_\alpha(L) = \arg \min_x E_P((\mathbf{1}_{\{x \geq L\}} - \alpha)(x - L))$$

using the strictly consistent scoring function $s(x, y) = (\mathbf{1}_{\{x \geq y\}} - \alpha)(x - y)$ from Example 2.7.

- (2) Taking this result as a fixed value \hat{q}_α , we observe that $E[L \mid L \geq \hat{q}_\alpha]$ is just an expected value. Thus, we can use a strictly consistent scoring function to forecast $ES_\alpha(L) \approx E[L \mid L \geq \hat{q}_\alpha]$. If L is square-integrable, the score function simply can be chosen as the squared error such that $ES_\alpha(L) \approx \arg \min_x E_{\tilde{P}}((x - L)^2)$, where $\tilde{P}(A) = P(A \mid L \geq \hat{q}_\alpha)$.

The result of this procedure is then a component-wise optimal forecast for the ES.

LEMMA 3.5 *For distributions with finite second moments, the variance is conditionally elicitable.*

PROOF OF LEMMA 3.5 Let $\mathcal{P} = \{\text{distributions on } \mathbb{R} \text{ with finite second moments}\}$. Defining γ_c by

$$\gamma_c: \mathcal{P} \rightarrow \mathbb{R}, \quad P \mapsto \gamma(P, c) := E_P[(L - c)^2]$$

and $\tilde{\gamma}$ by

$$\tilde{\gamma}: \mathcal{P} \rightarrow \mathbb{R}, \quad P \mapsto \tilde{\gamma}(P) := E_P(L),$$

we can rewrite the variance $E_P((L - E_P(L))^2)$ as $\text{var}(L) = \gamma(P, \tilde{\gamma}(P))$. Then $\tilde{\gamma}$ is elicitable according to Newey and Powell (1987). For fixed c , γ_c is elicitable according to Gneiting (2011, Theorem 8 (a)). It follows that the variance is conditionally elicitable in the sense of Definition 2.8. \square

LEMMA 3.6 (Gneiting 2011) *For distributions with finite means, expectiles are elicitable.*

As a consequence, by Remark 2.9, expectiles are conditionally elicitable.

4 CAPITAL ALLOCATION AND DIVERSIFICATION BENEFITS

For risk management purposes, it is useful to decompose the portfolio-wide risk into components (risk contributions) that are associated with the subportfolios or assets of which the portfolio is comprised. There are quite a few approaches to this problem (see Tasche (2008) for an overview). In the following, we discuss the so-called Euler allocation in more detail, as well as the quantification and comparison of the portfolio diversification.

4.1 Capital allocation using expected shortfall or expectiles

Tasche (1999) argues that, from an economic perspective with a view on portfolio optimization, it makes most sense to determine risk contributions as sensitivities (partial derivatives). What makes the definition of risk contributions by partial derivatives

even more attractive is the fact that by Euler's theorem (see Tasche (1999) for a statement of the theorem in a risk management context) such risk contributions add up to the portfolio-wide risk if the risk measure under consideration is homogeneous. Technically speaking, we suggest the following definition of risk contributions.

DEFINITION 4.1 Let L, L_1, \dots, L_m be random variables such that $L = \sum_{i=1}^m L_i$, and let ρ be a risk measure. If the derivative

$$\frac{d\rho(L + hL_i)}{dh}$$

exists for $h = 0$, then the risk contribution of L_i to $\rho(L)$ is defined by

$$\rho(L_i | L) = \left. \frac{d\rho(L + hL_i)}{dh} \right|_{h=0}. \quad (4.1)$$

If the derivatives on the right-hand side of (4.1) all exist for $i = 1, \dots, m$, and the risk measure ρ is homogeneous in the sense of Definition 2.1, then Euler's theorem implies

$$\rho(L) = \sum_{i=1}^m \rho(L_i | L).$$

Tasche (1999) shows that if one of the L_i has a smooth density conditional on the realizations of the other L_i , then the risk contributions of ES in the sense of Definition 4.1 all exist and have an intuitive shape. However, the process of identifying sufficient conditions for the existence of partial derivatives of a risk measure and their calculation can be tedious. For coherent risk measures, Delbaen (2000) described an elegant method to determine the risk contributions. In the following theorem, we present the risk contributions to ES. Then, in Theorem 4.3, we use the method in Delbaen (2000) to derive the risk contributions to expectiles.

THEOREM 4.2 (Tasche 1999; Delbaen 2000) *If the partial derivative as described in (4.1) exists for ρ chosen as the ES, then the risk contribution of a position L_i to the portfolio's ES can be calculated as*

$$\text{ES}_\alpha(L_i | L) = \mathbb{E}[L_i | L \geq q_\alpha(L)].$$

Using the approach in Delbaen (2000), we can also derive the capital allocation for expectiles. See Martin (2014) for an alternative approach based on saddlepoint approximation.

THEOREM 4.3 *If the partial derivative as described in (4.1) exists for $\rho = e_\tau$, then, for $\frac{1}{2} \leq \tau < 1$, the risk contribution of a position L_i to the portfolio's expectile can be calculated as*

$$e_\tau(L_i | L) = \frac{\tau \mathbb{E}[L_i \mathbf{1}_{\{L > e_\tau(L)\}}] + (1 - \tau) \mathbb{E}[L_i \mathbf{1}_{\{L \leq e_\tau(L)\}}]}{\tau P[L > e_\tau(L)] + (1 - \tau) P[L \leq e_\tau(L)]}. \quad (4.2)$$

PROOF OF THEOREM 4.3 The proof follows Delbaen's method (Delbaen 2000). Recall that the weak subgradient of a convex function $f: L^\infty(\Omega) \rightarrow \mathbb{R}$ at $X \in L^\infty(\Omega)$ (see Delbaen 2000, Section 8.1) is defined as

$$\nabla f(X) = \{\varphi: \varphi \in L^1(\Omega) \text{ such that for all } Y \in L^\infty(\Omega), \\ f(X + Y) \geq f(X) + \mathbb{E}[\varphi Y]\}.$$

In order to identify the subgradient of the risk measure e_τ , we note that

- e_τ is a law-invariant coherent risk measure;
- as shown in Jouini *et al* (2006), e_τ has the so-called Fatou property;
- as shown in Bellini *et al* (2014), we have that

$$e_\tau(L) = \max\{\mathbb{E}[\varphi L]: \varphi \in M_\tau\},$$

with

$$M_\tau = \left\{ \varphi \geq 0 \text{ is bounded with } \mathbb{E}[\varphi] = 1 \right. \\ \left. \text{and } \frac{\sup \varphi}{\inf \varphi} \leq \max\left(\frac{\tau}{1-\tau}, \frac{1-\tau}{\tau}\right) \right\};$$

- as shown in Bellini *et al* (2014), for

$$\bar{\varphi} = \frac{\tau \mathbf{1}_{\{L > e_\tau(L)\}} + (1-\tau) \mathbf{1}_{\{L \leq e_\tau(L)\}}}{\tau \mathbb{P}[L > e_\tau(L)] + (1-\tau) \mathbb{P}[L \leq e_\tau(L)]},$$

we have $\bar{\varphi} \in M_\tau$ and $e_\tau(L) = \mathbb{E}[\bar{\varphi} L]$.

Delbaen (2000, Theorem 17) now implies that $\bar{\varphi}$ is an element of $\nabla e_\tau(L)$, ie, it holds for all bounded random variables L^* that

$$e_\tau(L + L^*) \geq e_\tau(L) + \mathbb{E}[\bar{\varphi} L^*].$$

From Delbaen (2000, Proposition 5) it follows that, if $\nabla e_\tau(L)$ has only one element, then we have

$$\left. \frac{de_\tau(L + hL^*)}{dh} \right|_{h=0} = \mathbb{E}[\bar{\varphi} L^*]. \quad (4.3)$$

Taking $L^* = L_i$ in (4.3) implies (4.2). \square

The proof of Theorem 4.3 shows that risk contributions for expectiles (and also for ES) can be defined even if the derivatives in the sense of Definition 4.1 do not exist. This may happen if the distribution of the loss variable is not smooth (ie, not continuous). Then the subgradient set $\nabla e_\tau(L)$ may contain more than one element such that there is no unique candidate vector for the risk contributions (see Kalkbrener (2005) for more details on this approach to risk contributions for coherent risk measures).

4.2 Diversification benefits

In risk management, evaluating diversification benefits properly is key to both insurance and investments, since risk diversification may reduce a company's need for risk-based capital. To quantify and compare the diversification of portfolios, indexes have been defined, such as the closely related notions of diversification benefit, defined by Bürgi *et al* (2008), and the diversification index by Tasche (2008). These indexes are not universal risk measures and depend on the choice of risk measure and the number of underlying risks in the portfolio.

As mentioned earlier, the subadditivity and comonotonic additivity of a risk measure are important conditions for proper representation of diversification effects. In this case, capital allocation, as introduced in Section 4.1, can be helpful for identifying risk concentrations.

Let us define the diversification index (Tasche 2008).

DEFINITION 4.4 Let L_1, \dots, L_n be real-valued random variables, and let $L = \sum_{i=1}^n L_i$. If ρ is a risk measure such that $\rho(L), \rho(L_1), \dots, \rho(L_n)$ are defined, then

$$\text{DI}_\rho(L) = \frac{\rho(L)}{\sum_{i=1}^n \rho(L_i)}$$

denotes the diversification index of portfolio L with respect to the risk measure ρ . If risk contributions $\rho(L_i | L)$ of L_i to $\rho(L)$ (see Definition 4.1) exist, then

$$\text{DI}_\rho(L_i | L) = \frac{\rho(L_i | L)}{\rho(L_i)}$$

denotes the marginal diversification index of subportfolio L_i with respect to the risk measure ρ .

For the case of a homogeneous, subadditive and comonotonically additive risk measure, Tasche (2008) derived the following properties of the diversification index.

PROPERTIES 4.1 (Tasche 2008) Let ρ be a homogeneous, subadditive and comonotonically additive risk measure. Then,

- $\text{DI}_\rho(L) \leq 1$ (due to subadditivity);
- $\text{DI}_\rho(L) \approx 1$ indicates that L_1, \dots, L_n are “almost” comonotonic (the closer the index of diversification is to 1, the less diversified the portfolio);
- if $\text{DI}_\rho(L_i | L) < \text{DI}_\rho(L)$, then there exists $\varepsilon_i > 0$ such that $\text{DI}_\rho(L + hL_i) < \text{DI}_\rho(L)$ for all $0 < h < \varepsilon_i$.

It is not clear how far below 100% the diversification index should be to indicate high diversification, because, in the presence of undiversifiable risk, even a large

optimized portfolio might still have a relatively high index. Nonetheless, comparison between marginal diversification indexes and the portfolio's diversification index can be useful to detect unrealized diversification potential. Hence, instead of investigating the absolute diversification index, it might be better to look for high unrealized diversification potential as a criterion to judge a portfolio as highly concentrated.

Note that risk measures such as standard deviation or expectiles would show a 100% diversification index for portfolios with perfectly linearly correlated positions, but not for comonotonic positions with less than perfect linear correlation. Hence, for risk measures that are not comonotonically additive, there is a danger of underestimating the lack of diversification due to nonlinear dependence.

A notion similar to the diversification index was proposed in Bürgi *et al* (2008) to quantify the diversification performance of a portfolio of risks. Bürgi *et al* define the notion of diversification benefit, denoted by DB, of a portfolio $L = \sum_{i=1}^n L_i$ as

$$DB(L) = 1 - \frac{RAC_{\rho}(\sum_{i=1}^n L_i)}{\sum_{i=1}^n RAC_{\rho}(L_i)},$$

where RAC denotes the risk-adjusted capital, defined as the least amount of additional capital needed to prevent a company's insolvency at a given level of default probability:

$$RAC_{\rho}(L) = \rho(L) - E(L),$$

where $\rho(L)$ denotes the risk measure chosen for L . Clearly, DB has properties very similar to the properties of the the diversification index, namely the following.

PROPERTIES 4.2 (Bürgi *et al* 2008) Let ρ be a homogeneous, subadditive and comonotonically additive risk measure. Then,

- $0 \leq DB(L) \leq 1$ (due to subadditivity);
- the interpretation of the diversification benefit is straightforward, namely

$$DB(L) = \begin{cases} 1 & \text{indicates full hedging,} \\ 0 & \text{indicates comonotonic risks,} \\ x \in]0, 1[& \text{indicates that there is } 100x\% \text{ of capital} \\ & \text{reduction due to diversification,} \end{cases}$$

and hence, the higher $DB(L)$, the higher the diversification (in contrast to the diversification index DI_{ρ}).

The same comments apply to both Properties 4.1 and Properties 4.2. Both indexes depend not only on the choice of ρ and on the portfolio size n , but also, and even more strongly, on the dependence structure between the risks. Neglecting dependence may

lead to a gross underestimation of RAC. This has been analytically illustrated with a simple model in Busse *et al* (2014), where it was demonstrated that introducing dependence between the risks drastically reduces the diversification benefits.

When it comes to comparing the consequences of choosing VaR or ES, respectively, for the measurement of diversification benefits, we can really see the limitation of VaR as a risk measure. Even if there is a part of the risk that is undiversifiable, VaR might not catch it, as is demonstrated in Emmer and Tasche (2004, Proposition 3.3). In Busse *et al* (2014), VaR shows a diversification benefit for a very high number n of risks, while ES does not decrease for this range of n , thus correctly reflecting the fact that the risk cannot be completely diversified away.

Moreover, the type of dependence does matter. Linear dependence (measured with the linear correlation) cannot accurately describe dependence between extreme risks, particularly in times of stress. Neglecting the nonlinearity of dependence may lead to an overestimation of the diversification benefits. This is well described by Bürgi *et al* (2008), who consider elliptical and Archimedean copulas for risk modeling and compare their effect on the evaluation of RAC and, hence, the diversification benefit.

5 BACKTESTING: WHICH METHODS CAN BE USED?

What does backtesting mean? According to Jorion (2007), it is a set of statistical procedures designed to check if the realized losses, observed *ex post*, are in line with VaR forecasts. We may of course extend this definition to any risk measure.

Recently, Gneiting (2011) raised a potential issue with direct backtesting when using ES as a risk measure. This is not an issue for risk measures such as VaR or expectiles because of their elicibility, as seen previously. Is it a real issue in practice for ES? On the one hand, Acerbi and Székely (2014) have recently argued that, actually, elicibility (or a lack of elicibility) is not relevant for backtesting of risk measures but rather for comparing the forecast performance of different estimation methods. On the other hand, some financial institutions, in particular reinsurance companies, have addressed the problem of backtesting ES by using probability distribution forecasts for checking the output of their internal models. Nevertheless, if one still wants to stick to point forecasts only for ES, we propose an empirical approach that consists in approximating ES with quantiles (see Section 5.1).

Further, as observed in Section 3.3, ES is a combination of two elicitable components, since it is conditionally elicitable. A natural approach to the backtesting of ES therefore is to use the algorithm described in Section 3.3, where we backtest both components separately according to their associated respective scoring functions. Here, as the first component, we backtest the quantile. Then, taking the result for the quantile as a fixed value, we can backtest the ES, since it is then just a mean that has the quadratic error as its strictly consistent scoring function.

More generally, the choice of backtesting method should depend on the type of forecast. There are backtesting methods for the following.

- (i) Point forecasts for the value of a variable: they are usually represented as the conditional expectation $\mathbb{E}[Y_{t+k} \mid \mathcal{F}(Y_s, s \leq t)]$, where $\mathcal{F}(Y_s, s \leq t)$ represents the available information up to time t on the time series Y . There is a huge amount of literature, notably in econometrics, on point forecasts and on well-established methods for their out-of-sample backtesting (see, for example, Clements and Hendry 1998; Elliott *et al* 2006, 2013).
- (ii) Probability range forecasts or interval forecasts (eg, forecasts of VaR or ES): they project an interval in which the forecast value is expected to lie with some probability p (eg, the interval $(-\infty, \text{VaR}_p(Y_{t+k})]$, where $\text{VaR}_p(Y_{t+k})$ is the projected p -quantile of Y_{t+k}). Much work, in particular with regard to backtesting, has been done on interval forecasts in the last fifteen years. A good reference on this topic is Christoffersen (2003). Backtesting for VaR has been well developed, due to the interest of the financial industry in this risk measure. We refer, for example, to Davé and Stahl (1998) and, for a review on backtesting procedures for VaR, to Campbell (2006).
- (iii) Forecasts of the complete probability distribution $\mathbb{P}[Y_{t+k} \leq \cdot \mid \mathcal{F}(Y_s, s \leq t)]$ or its probability density function, if existing.

It is worth noting that, if there is a solution to (iii), there are also solutions for (i) and (ii). In addition, (iii) makes it possible to backtest ES, which helps us to avoid the issue raised by Gneiting (2011) for the direct backtesting of ES.

In contrast to VaR, ES is sensitive to the severity of losses exceeding the threshold VaR because the risk measure ES corresponds to the full tail of a distribution. Hence, seen as a part of the distribution beyond a threshold, the accuracy of the forecast of ES may be directly checked using tests on the accuracy of forecasts of probability distributions (see Tay and Wallis (2000) and Gneiting and Katzfuss (2014) for general discussions of this approach). Note that the tail of the distribution might be evaluated through a generalized Pareto distribution (GPD) above a high threshold via the Pickands theorem (see Embrechts *et al* 1997; Pickands 1974).

In the following, we provide more detail on (ii) and (iii).

5.1 Backtesting value-at-risk and expected shortfall

Backtesting value-at-risk

As mentioned in Example 2.7, VaR is elicited by the weighted absolute error scoring function (see Gneiting 2011; Saerens 2000; Thomson 1979), characterizing VaR as an optimal point forecast. This allows for the comparison of different forecast methods.

However, in practice, we have to compare VaR predictions using a single method with observed values to assess the quality of the predictions.

A popular procedure is based on the so-called violation process briefly described here. Since by definition of VaR, assuming a continuous loss distribution, we have $\mathbb{P}(L > \text{VaR}_\alpha(L)) = 1 - \alpha$, it follows that the probability of a violation of VaR is $1 - \alpha$. We define the violation process of VaR as

$$I_t(\alpha) = \mathbf{1}_{\{L(t) > \text{VaR}_\alpha(L(t))\}}.$$

Christoffersen (2003) showed that VaR forecasts are valid if and only if the violation process $I_t(\alpha)$ satisfies two conditions.

- The unconditional coverage hypothesis: $\mathbb{E}[I_t(\alpha)] = 1 - \alpha$.
- The independence condition: $I_t(\alpha)$ and $I_s(\alpha)$ are independent for $s \neq t$.

Under these two conditions, the $I_t(\alpha)$ are iid Bernoulli random variables with success probability $1 - \alpha$. Hence, the number of violations has a binomial distribution.

In practice, this means considering an estimate of the violation process by replacing VaR with its estimates and checking that this process behaves like iid Bernoulli random variables with violation (success) probability close to $1 - \alpha$. If the proportion of VaR violations is not significantly different from $1 - \alpha$, we conclude that the estimation/prediction method is reasonable.

However, the above independence condition might be violated in practice, such that the general way of computing VaR as an unconditional quantile from the historical sample seems questionable. This is why various tests on the independence assumption have been proposed in the literature, for example, that developed by Christoffersen and Pelletier (2004) based on the duration of days between the violations of the VaR thresholds.

Backtesting expected shortfall

A similarly simple approximative approach to the backtesting of ES might be based on a representation of ES as integrated VaR (Acerbi and Tasche 2002, Proposition 3.2):

$$\begin{aligned} \text{ES}_\alpha(L) &= \frac{1}{1 - \alpha} \int_\alpha^1 q_u(L) \, du \\ &\approx \frac{1}{4} [q_\alpha(L) + q_{0.75\alpha+0.25}(L) + q_{0.5\alpha+0.5}(L) + q_{0.25\alpha+0.75}(L)], \end{aligned} \quad (5.1)$$

where $q_\alpha(L) = \text{VaR}_\alpha(L)$. Hence, if $q_\alpha(L)$, $q_{0.75\alpha+0.25}(L)$, $q_{0.5\alpha+0.5}(L)$ and $q_{0.25\alpha+0.75}(L)$ are successfully backtested, then the estimate of $\text{ES}_\alpha(L)$ can be considered reliable, subject to a careful manual inspection of the observations in the upper 0.25% tail of the observed sample. The upper tail observations must be manually

inspected anyway in order to separate data outliers from genuine far tail observations. Hence, the suggested procedure provides a reasonable combination of statistical testing and human oversight. Compared with the test procedures suggested in Acerbi and Székely (2014), it has the advantage of not relying on Monte Carlo simulation for the statistical test.

Do four supporting points suffice in the linear approximation to ES by different VaRs in (5.1)? As a matter of fact, the power of the joint test for VaR violations on the supporting points will decline with the number of supporting points chosen; however, it will increase with the size of the sample of available observations. Hence, the number of supporting points must be determined on a case-by-case basis with a view on the sample size.

The approach based on (5.1) is attractive not only because of its simplicity but also because it illustrates the fact that, for the same level of certainty, a much longer sample is needed for the validation of $ES_{\alpha}(L)$ than for $VaR_{\alpha}(L)$ (see also Yamai and Yoshihara 2005). The Basel Committee suggests a variant of this ES-backtesting approach that is based on testing level violations for two quantiles at the 97.5% and 99% levels (Basel Committee on Banking Supervision 2013).

5.2 Backtesting distribution forecasts

Let us outline a method for the out-of-sample validation of distribution forecasts, based on the Lévy–Rosenblatt transform, which is also called the probability integral transform (PIT). As pointed out before, this methodology is important since testing the distribution forecasts could be helpful, in particular, for tail-based risk measures such as ES.

The use of the PIT for backtesting financial models is relatively recent. The foundations were laid by Diebold *et al* (1998). These authors tackled the problem of density forecast evaluation from a risk management perspective, suggesting a method for testing continuous distribution forecasts in finance based on the uniform distribution of the Lévy–Rosenblatt transform (or PIT) (Lévy 1937; Rosenblatt 1952). Applying the Lévy theorem to the PIT, Diebold *et al* (1998) observed that if a sequence of distribution forecasts coincides with the sequence of unknown conditional laws that have generated the observations, the sequence of the PIT is independent and identically $\mathcal{U}(0, 1)$ -distributed. In Diebold *et al* (1999), they extended the density forecast evaluation to the multivariate case, involving cross-variable interactions such as time-varying conditional correlations, and provided conditions under which a technique of density forecast “calibration” can be used to improve deficient density forecasts. They finally applied the PIT method on high-frequency financial data (volatility forecasts) to illustrate its application. Note that the definition of PIT has been generalized

for not necessarily continuous cumulative distribution functions (see Gneiting and Ranjan (2013) and the references therein).

Nevertheless, there was still a gap left to fill before a full implementation and use in practice. In his PhD thesis, Blum (2004) studied various issues that had been left open and proposed and validated mathematically a method based on the PIT in situations with overlapping forecast intervals and multiple forecast horizons. Blum (2004) illustrated this in his thesis dealing with economic scenario generators (ESG). Typically, financial institutions make use of scenario generators, producing thousands of scenarios, each with its own forecast value for a certain value at a certain future time. Recall that the scenarios are constructed by simulating the iid innovations of the underlying process. Those simulated values define an empirical distribution, which represents a distribution forecast. Hence, the backtesting will be done on the obtained distribution; it is an out-of-sample backtesting of distribution forecasts. For details of the methodology, we refer to Blum (2004), SCOR Switzerland AG (2008) and the references therein, and only summarize the main steps in the following.

Using the values obtained from all the scenarios, we deduce the empirical distribution denoted by $\hat{\Phi}_i$, which is assumed to converge to the marginal cumulative distribution function Φ_i defined by $\Phi_i(x) = \mathbb{P}(X_i \leq x \mid \mathcal{F}_{i-m})$, where X_i corresponds to the scenario forecast of a variable X at out-of-sample time point t_i and \mathcal{F}_{i-m} corresponds to the information available up to time t_{i-m} from the simulation start, with m being the number of forecast steps. Hence, at out-of-sample time point t_i , we make use of $\hat{\Phi}_i$ when identifying the distribution at time t_i as the one computed at the previous time t_{i-m} , and a newly observed value x_i .

Now we apply the PIT to build the random variables $Z_i := \hat{\Phi}_i(X_i)$, with known realizations $\hat{\Phi}_i(x_i)$. These have been proved by Diebold *et al* (1998, 1999) to be independent and identically $\mathcal{U}(0, 1)$ -distributed whenever the conditional distribution forecast $\Phi_i(\cdot)$ coincides with the true process by which the historical data has been generated.

For practical purposes, it then suffices to test if the PIT-transformed variables Z_i are independent and identically $\mathcal{U}(0, 1)$ -distributed. If one of these conditions is rejected, the model does not pass the out-of-sample test. As noted by Diebold and Mariano (1995), this is not a test on the model, so it does not mean the model is valueless. Rejection only means that there may be a structural difference between the in-sample and out-of-sample periods, or that the model does not hold up to the full predictive data.

Various statistical tests are possible, like standard tests such as the χ^2 test for uniformity or the Kendall–Stuart test for the significance of the autocorrelations. If we are continuing with the Diebold *et al* methodology, the nonparametric test proposed in Diebold *et al* (1998) (see also Diebold *et al* (1999) for the multivariate case) may also be useful. This test consists of comparing histograms obtained from Z_i

and $\mathcal{U}(0, 1)$, respectively, and detecting deviations from the independence property when considering correlograms of the Z_i and their lower integer powers.

Note that tests based on the PIT have some limitation due to serial correlation. One way of overcoming this issue is, for example, as suggested in SCOR Switzerland AG (2008), generating realistic forecast scenarios via refined bootstrapping.

Many other results have enriched the literature on distribution backtesting (see, for example, Elliott *et al* 2006, 2013; Gneiting and Katzfuss 2014). We mention two other methods to complete our review: one is based on the notion of scoring (see, for example, Gneiting and Raftery (2007), Gneiting and Ranjan (2011), Amisano *et al* (2007) or the survey paper of Gneiting and Katzfuss (2014)); the other mixes the scoring and PIT approaches (see Gneiting and Katzfuss 2014). We already introduced the concept of scoring function s in Definition 2.4. When using it for backtesting purposes, we modify it to measure the loss function $s(f, Y)$, whose arguments are the density forecast f and the realization y of the future observation Y .

6 CONCLUSION

In this paper, we have listed a number of properties that are commonly considered to be must-haves for good risk measures: coherence, comonotonic additivity, robustness and elicibility. We have then revisited the popular risk measures VaR and ES as well as the recently suggested expectiles and checked which of the following properties they satisfy.

- It is well known that VaR lacks subadditivity in general and, therefore, might fail to appropriately account for risk concentrations. However, we found that for many practical applications this might not be a serious issue, as long as the underlying risks have a finite variance or, in some cases, a finite mean. The fact that VaR does not cover tail risks “beyond” VaR is a more serious deficiency, although ironically it makes VaR a risk measure that is more robust than the other risk measures we have considered. This deficiency can be particularly serious when one faces choices of various risks with different tails. VaR and ES will present different optimal results that are known to be suboptimal in terms of risk for VaR (see, for example, McNeil *et al* 2005, Example 6.7).
- ES makes good the lack of subadditivity and sensitivity to tail risk in VaR, but it has recently been found to be not elicitable. This means that the backtesting of ES is less straightforward than the backtesting of VaR. We have found that, nonetheless, there are a number of feasible approaches to the backtesting of ES (eg, based on distribution forecasts, the linear approximation of ES with VaR at different confidence levels or directly with Monte Carlo tests). However, it

must be conceded that, to reach the same level of certainty, more validation data is required for ES than for VaR.

- Expectiles have been suggested as coherent and elicitable alternatives to ES. However, while expectiles do indeed have a number of attractive features, their underlying concept is less intuitive than the concepts for VaR or ES. In addition, expectiles are not comonotonically additive, which implies that in applications they may fail to detect risk concentrations due to nonlinear dependencies.

To conclude, we have found that among the risk measures we discussed, ES seems to be the best for use in practice, despite some caveats with regard to its estimation and backtesting, which can be carefully mitigated. We have not found sufficient evidence to justify an all-inclusive replacement of ES by its recent competitor expectile. Nonetheless, it is certainly worthwhile to keep expectiles in mind as an alternative to ES and VaR in specific applications.

DECLARATION OF INTEREST

The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Bank of England. Partial support from RARE-318984 (an FP7 Marie Curie IRSES Fellowship) is kindly acknowledged.

ACKNOWLEDGEMENTS

We thank two referees for their careful reading of the manuscript and suggestions that helped to improve the presentation of the paper. We are also grateful to Michel Dacorogna for interesting discussions on backtesting in practice.

REFERENCES

- Acerbi, C., and Székely, B. (2014). Back-testing expected shortfall. *Risk* **27**(11), 1–6.
- Acerbi, C., and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance* **26**, 1487–1503.
- Amisano, G., and Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics* **25**, 177–190.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risks. *Mathematical Finance* **9**, 203–228.
- Basel Committee on Banking Supervision (2013). Fundamental review of the trading book: a revised market risk framework. Consultative Document, October, Bank for International Settlements.
- Bellini, F., and Bigozzi, V. (2015). On elicitable risk measures. *Quantitative Finance* **15**(5), 725–733.

- Bellini, F., Klar, B., Müller, A., and Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics* **54**, 41–48.
- Blum, P. (2004). On some mathematical aspects of dynamic financial analysis. PhD Thesis, ETH Zürich.
- Bürgi, R., Dacorogna, M. M., and Iles, R. (2008). Risk aggregation, dependence structure and diversification benefit. In *Stress Testing for Financial Institutions*, pp. 265–306. Risk Books, London.
- Busse, M., Dacorogna, M. M., and Kratz, M. (2014). The impact of systemic risk on the diversification benefits of a risk portfolio. *Risks* **2**(3), 260–276.
- Campbell, S. D. (2006). A review of backtesting and backtesting procedures. *Journal of Risk* **9**(2), 1–17.
- Chen, J. M. (2014). Measuring market risk under the Basel Accords: VaR, stressed VaR, and expected shortfall. *Aestimatio: IEB International Journal of Finance* **8**, 184–201.
- Christoffersen, P. (2003). *Elements of Financial Risk Management*. Academic Press.
- Christoffersen, P., and Pelletier, D. (2004). Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics* **2**(1), 84–108.
- Clements, M., and Hendry, D. (1998). *Forecasting Economic Time Series*. Cambridge University Press.
- Cont, R., Deguest, R., and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance* **10**(6), 593–606.
- Cramér, H. (1930). *On the Mathematical Theory of Risk*. Skandia Jubilee Volume, Stockholm.
- Danielson, J., Jorgenson, B., Samorodnitsky, G., Sarma, M., and de Vries, C. (2013). Fat tails, VaR and subadditivity. *Journal of Econometrics* **172**(2), 283–291.
- Davé, R. D., and Stahl, G. (1998). On the accuracy of VaR estimates based on the variance–covariance approach. In *Risk Measurement, Econometrics and Neural Networks*, Bol, G., Nakhaeizadeh, G., and Vollmer, K.-H. (eds), pp. 189–232. Springer.
- Delbaen, F. (2000). Coherent risk measures. PISA Lecture Notes, ETH Zürich. URL: www.math.ethz.ch/~delbaen/ftp/preprints/PISA007.pdf.
- Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**, 253–263.
- Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with application to financial risk management. *International Economic Review* **39**, 863–883.
- Diebold, F. X., Hahn, J., and Tay, A. S. (1999). Multivariate density forecasts and calibration in financial risk management: high-frequency returns on foreign exchange. *Review of Economics and Statistics* **81**, 661–673.
- Elliott, G., Granger, C., and Timmermann, A. (2006). *Handbook of Economic Forecasting*, Volume 1. Elsevier.
- Elliott G., Granger, C., and Timmermann, A. (2013). *Handbook of Economic Forecasting*, Volume 2. Elsevier.
- Embrechts, P., and Hofert, M. (2014). Statistics and quantitative risk management for banking and insurance. *Annual Review of Statistics and Its Applications* **1**, 493–514.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.

- Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, Dempster, M., and Howarth, A. (eds), pp. 176–223. Cambridge University Press.
- Embrechts, P., Lambrigger, D., and Wüthrich, M. (2009a). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes* **12**(2), 107–127.
- Embrechts, P., Neslehová, J., and Wüthrich, M. (2009b). Additivity properties for value-at-risk under Archimedean dependence and heavy-tailedness. *Insurance: Mathematics and Economics* **44**, 164–169.
- Embrechts, P., Puccetti, G., and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance* **37**(8), 2750–2764.
- Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks* **2**(1), 25–48.
- Emmer, S., and Tasche, D. (2004). Calculating credit risk capital charges with the one-factor model. *Journal of Risk* **7**(2), 85–103.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association* **106**(494), 746–762.
- Gneiting, T., and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and its Applications* **1**, 125–151.
- Gneiting, T., and Raftery, A. E. (2007). Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association* **102**(477), 359–378.
- Gneiting, T., and Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted proper scoring rules. *Journal of Business and Economic Statistics* **29**, 411–422.
- Gneiting, T., and Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics* **7**, 1747–1782.
- Hauksson, A., Dacorogna, M. M., Domenig, T., Müller, U. A., and Samorodnitsky, G. (2001). Multivariate extremes, aggregation and risk estimation. *Quantitative Finance* **1**, 79–95.
- Huber, P. J., and Ronchetti, E. M. (2009). *Robust Statistics*, 2nd edn. Wiley.
- Jadhav, D., Ramanathan, T. V., and Naik-Nimbalkar, U. V. (2013). Modified expected shortfall: a new robust coherent risk measure. *Journal of Risk* **16**(1), 69–83.
- Jorion, P. (2007). *Financial Risk Manager Handbook*, 4th edn. Wiley.
- Jouini, E., Schachermayer, W., and Touzi, N. (2006). Law invariant risk measures have the Fatou property. In *Advances in Mathematical Economics*, pp. 49–71. Springer.
- Kalkbrener, M. (2005). An axiomatic approach to capital allocation. *Mathematical Finance* **15**(3), 425–437.
- Kratz, M. (2013). There is a VaR beyond usual approximations. Towards a toolkit to compute risk measures of aggregated heavy tailed risks. Report, Swiss Financial Market Supervisory Authority, FINMA.
- Kratz, M. (2014). Normex, a new method for evaluating the distribution of aggregated heavy tailed risks: application to risk measures. *Extremes, Special Issue: Extremes and Finance* **17**(4), 661–691.
- Lambert, N. S., Pennock, D. M., and Shoham, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pp. 129–138. Association for Computing Machinery.

- Lévy, P. (1937). *Théorie de l'Addition des Variables Aléatoires*. Gauthier-Villars, Paris.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance* **7**(1), 77–91.
- Martin, R. J. (2014). Expectiles behave as expected. *Risk* **27**(3), 79–83.
- McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management*. Princeton University Press.
- Newey, W., and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55**(4), 819–847.
- Osband, K. H. (1985). Providing incentives for better cost forecasting. PhD Thesis, University of California, Berkeley.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119–131.
- Rockafellar, R. T., and Uryasev, S. (2013). The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science* **18**(1), 33–53.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics* **23**, 470–472.
- Saerens, M. (2000). Building cost functions minimizing to some summary statistics. *IEEE Transactions on Neural Networks* **11**, 1263–1271.
- SCOR Switzerland AG (2008). *From Principle-Based Risk Management to Solvency Requirements: Analytical Framework for the Swiss Solvency Test*. SCOR Switzerland AG.
- Stahl, G., Zheng, J., Kiesel, R., and Rühlicke, R. (2012). Conceptualizing robustness in risk management. *SSRN Electronic Journal*, DOI: 10.2139/ssrn.2065723.
- Tasche, D. (1999). Risk contributions and performance measurement. Working Paper, Technische Universität München.
- Tasche, D. (2002). Expected shortfall and beyond. *Journal of Banking and Finance* **26**(7), 1519–1533.
- Tasche, D. (2008). Capital allocation to business units and sub-portfolios: the Euler principle. In *Pillar II in the New Basel Accord: The Challenge of Economic Capital*, Resti, A. (ed), pp. 423–453. Risk Books, London.
- Tay, A., and Wallis, K. (2000). Density forecasting: a survey. *Journal of Forecasting* **19**, 235–254.
- Thomson, W. (1979). Eliciting production possibilities from a well-informed manager. *Journal of Economic Theory* **20**, 360–380.
- Yamai, Y., and Yoshiba, T. (2005). Value-at-risk versus expected shortfall: a practical perspective. *Journal of Banking and Finance* **29**(4), 997–1015.
- Ziegel, J. F. (2014). Coherence and elicibility. *Mathematical Finance*, DOI: 10.1111/mafi.12080.