What Is the Best Risk Measure in Practice? A Comparison of Standard Measures

Research Center
ESSEC Working Paper 1322

2013

Susanne Emmer
Marie Kratz
Dirk Tasche

ESSEC
BUSINESS SCHOOL

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

Susanne Emmer [1]*, Marie Kratz [2]**, Dirk Tasche[3,†]

[1] CREAR, ESSEC Business School Paris
[2] ESSEC Business School, CREAR
[3] E-mail: dirk.tasche@gmx.net

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 criticised for issues relating to backtesting. In particular, ES has been found not to be elicitable which means that backtesting for ES is less straight-forward than, e.g., backtesting for VaR. Expectiles have been suggested as potentially better alternatives to both ES and VaR. In this paper, we revisit commonly accepted desirable properties of risk measures like coherence, comonotonic additivity, robustness and elicitability. 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. In particular, there is no sufficient evidence to justify an all-inclusive replacement of ES by Expectiles in applications, especially as we provide an alternative way for backtesting of ES.
2000 AMS classification: 62P05; 91B30

Keywords: Backtesting; capital allocation; coherence; diversification; elicitability; expected shortfall; expectile; forecasts; probability integral transform (PIT); risk measure; risk management; robustness; value-at-risk

* Susanne Emmer has been visiting professor at ESSEC Business School in 2012-2013 and is an associated member of CREAR risk research center.
** Marie Kratz is also member of MAP5, UMR 8145, Univ. Paris Descartes.
† Dirk Tasche currently works at the Prudential Regulation Authority (a division of the Bank of England). The opinions expressed in this paper are those of the authors and do not necessarily reflect views of the Bank of England.

1 Introduction

Risk Management is a core competence of financial institutions like banks, insurance companies, investment funds and others. Techniques for the measurement of risk are clearly central for 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 there exist a 'best' risk measure?

Much research in economics, finance, and mathematics has been devoted to answer those questions. Cramér (1930) was one of the earliest researchers on risk capital, introducing ruin theory ([14]). A major contribution was made by Markowitz (1952, [39]) with modern portfolio theory. The variance of the Profit and Loss (P&L) distribution became then 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. in 1999 ([2]) desirable properties of risk measures have been formalized in a set of axioms. Because Expected Shortfall 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 ([46]). The Basel Committee on Banking Supervision also recommends replacing VaR by ES in internal market risk models [3]. Recently, a study by Gneiting ([28]) has pointed out that there is an issue with direct backtesting of Expected Shortfall estimates because Expected Shortfall is not *elicitable*. Therefore, with a view on the feasibility of backtesting, in recent studies ([4, 48]) Expectiles have been suggested as coherent and elicitable alternatives to Expected Shortfall. See also [10] for a detailed discussion of the issue.

In this paper, we discuss and compare the properties of some popular risk measures based on the loss distribution in order to provide answers to the questions raised above and to study their impact in terms of risk management.

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 $(\Omega, \mathcal{F}, P)$.

The paper is organized as follows: After the introductory section 1, Section 2 recalls the main definitions and properties of what is expected from a risk measure, like coherence,

comonotonic additivity, law invariance, elicitability 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 subadditivity of Value-at-Risk, we look at different concepts of robustness, discuss the elicitability of Expected Shortfall and Expectiles, and prove that Expectiles are not comonotonically additive. Section 4 deals with capital allocation and diversification benefits, 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. Furthermore, we introduce the concept of diversification index for the quantification and comparison of the diversification of portfolios. We then present in Section 5 methods for backtesting in general and look in more detail at Expected Shortfall. The paper ends in Section 6 with a discussion of the advantages and disadvantages of the different risk measures and a recommendation for the choice of a risk measure in practice.

2 Risk measures: definition and basic properties

Risk and risk measure are terms that have no unique definition and usage. It would be natural to measure risk in terms of probability distributions. But often it is useful to express risk with one number. Mappings from spaces of probability distributions or random variables into the real numbers are called risk measures. In this paper, a risk measure is understood as providing a risk assessment in 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. [2] 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. 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. 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. call the investor's future net worth 'risk'. Moreover, they state four axioms which 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*

$$\rho(h\,L) = h\,\rho(L). \tag{2.1}$$

[1] See [40] for alternative interpretations of risk measures.

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

$$\rho(L_1 + L_2) \;\leq\; \rho(L_1) + \rho(L_2). \tag{2.2}$$

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

$$L_1 \leq L_2 \;\Rightarrow\; \rho(L_1) \leq \rho(L_2). \tag{2.3}$$

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

$$\rho(L - a) = \rho(L) - a. \tag{2.4}$$

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

DEFINITION 2.2 *Two real-valued random variables L_1 and L_2 are said* comonotonic *if there exist a real-valued random variable X (the common risk factor) and non-descreasing functions f_1 and f_2 such that*

$$L_1 = f_1(X) \quad \textit{and} \quad L_2 = f_2(X). \tag{2.5a}$$

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

$$\rho(L_1 + L_2) \;=\; \rho(L_1) + \rho(L_2). \tag{2.5b}$$

Comonotonicity may be considered the strongest possible dependence of random variables ([24]). 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 does not attribute any diversification benefits to comonotonic risks (via comonotonic additivity) – which appears quite intuitive. Risk measures that depend only on the distributions of the losses are of special interest because their values can be estimated from loss observations only (i.e. no additional information like stress scenarios is needed).

DEFINITION 2.3 *A risk measure ρ is* ***law-invariant*** *if*

$$P(L_1 \leq \ell) = P(L_2 \leq \ell),\ \ell \in \mathbb{R} \;\Rightarrow\; \rho(L_1) = \rho(L_2). \tag{2.6}$$

2.2 Elicitability

An interesting criterion when estimating and backtesting a risk measure is elicitability ([28]). For the definition of elicitability we first introduce the concept of strictly consistent scoring functions.

DEFINITION 2.4 *A* ***scoring function*** *is a function*

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

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

EXAMPLE 2.1 *The following examples of score functions are of interest for this paper:*

$$\begin{aligned} s(x,y) &= (x-y)^2, \text{ squared error} \\ s(x,y) &= \begin{cases} \tau\,(x-y)^2, & x < y \\ (1-\tau)\,(y-x)^2, & x \geq y \end{cases}, \ 0 < \tau < 1 \text{ fixed, weighted squared error} \\ s(x,y) &= |x-y|, \text{ absolute error} \\ s(x,y) &= \begin{cases} \alpha\,(x-y), & x < y \\ (1-\alpha)\,(y-x), & x \geq y \end{cases}, \ 0 < \alpha < 1 \text{ fixed, weighted absolute error} \end{aligned}$$

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

$$\nu : \mathcal{P} \to 2^{\mathbb{R}}, \ \ P \mapsto \nu(P) \subset \mathbb{R},$$

*where $2^{\mathbb{R}}$ denotes the power set of $\mathbb{R}$. A scoring function $s : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ 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 $x \in \mathbb{R}$,*

$$\mathrm{E}_P\left[s(t,L)\right] \ \leq \ \mathrm{E}_P\left[s(x,L)\right].$$

*The function s is **strictly consistent** if it is consistent and*

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

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

Elicitability is a helpful criterion for the determination of optimal 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

$$\hat{x} = \arg\min_x \mathrm{E}_P\left[s(x,L)\right]$$

Hence elicitability of a functional of probability distributions may be interpreted as the property that the functional can be estimated by generalised regression. So far we have only distinguished between elicitable and non-elicitable functionals. However, it turns out that some useful risk measures are not elicitable but almost elicitable in the following sense.

DEFINITION 2.7 (CONDITIONAL ELICITABILITY) *A functional ν of $\mathcal{P}$ is called **conditionally elicitable** if there exist functionals $\widetilde{\gamma}$ and γ such that*

$$\nu(P) = \gamma(P, \widetilde{\gamma}(P)),$$

where $\widetilde{\gamma}$ is elicitable relative to $\mathcal{P}$ and γ is such that γ_c defined by

$$\gamma_c : \mathcal{P} \to 2^{\mathbb{R}}, \ P \mapsto \gamma(P,c) \subset \mathbb{R}$$

is elicitable relative to $\mathcal{P}$ for all $c \in \mathbb{R}$.

2.3 Robustness

Another important issue when estimating risk measures is robustness. Without robustness, results are not meaningful, since then 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 which are convenient in risk management, one usually considers the Wasserstein distance when investigating the robustness of risk measures.

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

$$d_W(P,Q) = \inf\{E[|X-Y|] : X \sim P, Y \sim Q\} \tag{2.7}$$

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 2.8 ([4]) *Let P_n, $n \geq 1$, and P be probability measures, and $X_n \sim P_n$, $n \geq 1$ and $P \sim X$. A risk measure ρ is called continuous at X with respect to the Wasserstein distance if*

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

Cont et al. ([13]) use a slightly different, potentially more intuitive concept of robustness which 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 basis for estimation. It turns out that for the same risk measure the estimation method can have a significant impact on the sensitivity. For instance, 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.

Cont et al. 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. allows to distinguish between different degrees of robustness. However, this concept makes 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. That is why for the purpose of this paper we adopt a notion of robustness based on the Wasserstein distance.

2.4 Popular risk measures

Variance and standard deviation were historically the dominating risk measures in finance. However, in the past 20 years or so, they have often been replaced in practical applications by VaR, which is currently the most popular downside risk measure.

DEFINITION 2.9 *The **Value-at-Risk (VaR)** at level $\alpha \in (0,1)$ of a loss variable L is defined as the α-quantile of the loss distribution:*

$$\mathrm{VaR}_\alpha(L) = q_\alpha(L) = \inf\{\ell :\ P(L \leq \ell) \geq \alpha\}. \tag{2.8}$$

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 Expected Shortfall was introduced to solve these issues.

DEFINITION 2.10 ([1]) *The **Expected Shortfall (ES)** at level $\alpha \in (0,1)$ (also called Tail Value-at-Risk or Superquantile) of a loss variable L is defined as*

$$\begin{aligned} \mathrm{ES}_\alpha(L) &= \frac{1}{1-\alpha}\int_\alpha^1 q_u(L)du \\ &= \mathrm{E}[L|L \geq q_\alpha(L)] + (\mathrm{E}[L|L \geq q_\alpha(L)] - q_\alpha(L))\left(\frac{\mathrm{P}[L \geq q_\alpha(L)]}{1-\alpha} - 1\right). \end{aligned} \tag{2.9}$$

If $\mathrm{P}[L = q_\alpha(L)] = 0$ *(in particular, if L is continuous),* $\mathrm{ES}_\alpha(L) = \mathrm{E}[L|L \geq q_\alpha(L)]$.

ES has been shown not to be elicitable ([28]). That is why Expectiles have been suggested as coherent and elicitable alternatives ([4, 48]).

DEFINITION 2.11 *For $0 < \tau < 1$ and square-integrable L, the τ-**Expectile** $e_\tau(L)$ is defined as*

$$e_\tau(L) = \arg\min_{\ell \in \mathbb{R}} \mathrm{E}[\tau \max(L-\ell, 0)^2 + (1-\tau)\max(\ell - L, 0)^2] \tag{2.10}$$

Note that, as for the variance, the notion of Expectile requires finite second moments.

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 one figure, which should be reported as a basis for strategic decisions. To help for this choice, let us start by giving an overview over the considered risk measures and their properties (see Table 1), before coming back to them with more details.

3.1 Coherence

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, VaR at level α gives no information about the severity of tail losses which occur with a probability less than $1-\alpha$, in contrast to ES at the same confidence level.

[2] According to [43] 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.

Table 1: Properties of standard risk measures

Property	variance	VaR	ES	e_τ(for $\tau \geq 1/2$)
Coherence			x	x
Comonotonic additivity		x	x	
Robustness w.r.t. weak topology		x[2]		
Robustness w.r.t. Wasserstein distance	x	x	x	x
Elicitability		x		x
Conditional Elicitability	x	x	x	x

When looking at aggregated risks $\sum_{i=1}^n L_i$, it is well known ([1]) that the risk measure ES is coherent. In particular it is subadditive, i.e.

$$ES_\alpha\big(\sum_{i=1}^n L_i\big) \;\leq\; \sum_{i=1}^n ES_\alpha(L_i).$$

In contrast, VaR is not subadditive in general. Indeed, examples (see e.g. [23]) can be given where it is superadditive, i.e.

$$VaR_\alpha\big(\sum_{i=1}^n L_i\big) > \sum_{i=1}^n VaR_\alpha(L_i).$$

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, but 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 e.g. [15], [23], [24], [25] or [26].

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

PROPOSITION 3.1 *([23]) Consider independent and identically distributed random variables X_i, $i = 1, \ldots, n$ with parent random variable X and cumulative distribution function F_X.*

Assume they are regularly varying with tail index $\beta > 0$, which means that the right tail $1 - F_X$ of their distribution satisfies

$$\lim_{x\to\infty} \frac{1 - F_X(ax)}{1 - F_X(x)} = a^{-\beta}, \text{ for all } a > 0$$

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

$$\lim_{\alpha\nearrow 1} \frac{VaR_\alpha\big(\sum_{i=1}^n X_i\big)}{\sum_{i=1}^n VaR_\alpha(X_i)} \leq 1 \quad \Leftrightarrow \quad \beta \geq 1.$$

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

PROPOSITION 3.2 ([24]) *Let $X = (X_1, \ldots, X_n)$ be a random vector having an elliptical distribution. Consider the set of linear portfolios $M = \{Z = \sum_{i=1}^n \lambda_i X_i \mid \sum_{i=1}^n \lambda_i = 1\}$.*
Then VaR at level α is subadditive on M if $0.5 < \alpha < 1$:

$$VaR_\alpha(Z_1 + Z_2) \leq VaR_\alpha(Z_1) + VaR_\alpha(Z_2), \quad Z_1, Z_2 \in M.$$

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

PROPOSITION 3.3 ([25]) *Consider random variables X_i, $i = 1, \ldots, n$ which have the same continuous marginal distribution function F. Assume the tail distribution $\bar{F} = 1 - F$ is regularly varying with tail index $-\beta < 0$, i.e. $\bar{F}(x) = x^{-\beta}G(x)$ for some function G slowly varying at infinity, and assume $(-X_1 \ldots, -X_n)$ has an Archimedean copula with generator Ψ, which is regulary 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 non-coherent measures such as VaR. In this paper, we do not review all these techniques and results but refer to [26] and the references therein for an overview.

Nevertheless, it is worth mentioning two recent studies, a new numerical algorithm introduced by Embrechts and co-authors ([26]) to provide bounds of VaR of aggregated risks, and a study by Kratz ([32], [33]) on the evaluation of VaR of aggregated heavy tailed risks. The numerical algorithm introduced in [26] allows for the computation of reliable lower and upper bounds for the VaR of high-dimensional (inhomogeneous) portfolios, whatever the dependence structure is. Quoting the authors, "surprisingly, additional positive dependence information (like positive correlation) does typically not improve the upper bound substantially. In contrast higher order marginal information on the model, when available, may lead

to strongly improved bounds. It 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 [33], a new approach, called Normex, is developed to provide accurate estimates of high quantiles for aggregated heavy tailed risks. This method depends only weakly upon the sample size and gives good results for any non-negative 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 Proposition 2.1 of [43], in risk management one usually considers robustness as continuity with respect to the Wasserstein distance as defined by (2.7). According to Stahl et al. ([43]), variance, Expected Shortfall, 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. observe that mean, VaR, and Expected Shortfall are continuous with respect to the Wasserstein distance and Bellini et al. ([4]) show that Expectiles are Lipschitz-continuous with respect to the Wasserstein distance with constant $K = \max\{\frac{\alpha}{1-\alpha}; \frac{1-\alpha}{\alpha}\}$, which implies continuity with respect to the Wasserstein distance.

With regard to robustness in the sense given in [13] (as mentioned in section 2.3), Cont et al. demonstrate that historical Expected Shortfall 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 Expected Shortfall at 99% level is much more sensitive than Gaussian and Laplace Expected Shortfall. 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. are not too surprising. The notion of ES was introduced precisely as a remedy to the lack of risk sensitivity of VaR.

Finally, note that in practice, the estimation of ES will be based on larger subsamples than the estimation of VaR. For instance, when using $100,000$ simulation iterations, ES at 99% level is estimated with 1,000 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 VaR and ES of aggregated financial returns, Hauksson et al. ([29]) noticed that the numerical stability of the scaling exponent was much higher with ES. This observation, in a way, counters the comments of Cont 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 Conditional Elicitability of Expected Shortfall

The lack of coherence of Value-at-Risk (VaR), which is up to now the most popular risk measure in practice, draws the attention to another downside risk measure, Expected Shortfall

(ES) as defined in (2.9). Expected Shortfall is a coherent risk measure and, in contrast to Value-at-Risk, is sensitive to the severity of losses beyond Value-at-Risk. Nevertheless, as soon as it comes to forecasting and backtesting Expected Shortfall, a potential deficiency arises compared to Value-at-Risk. Gneiting [28]) showed that Expected Shortfall is not elicitable. He proved that the existence of convex level sets is a necessary condition for the elicitability of a risk measure and disproved the existence of convex level sets for the Expected Shortfall. It is interesting to note that other important risk measures like the variance are not elicitable either ([36]).

Expected Shortfall is not elicitable, but, like the variance, is conditionally elicitable. This is a straightforward application of Definition 2.7 of conditional elicitability, noticing that Expected Shortfall can be represented as a combination of $E[L|L \geq c]$ and $c = q_\alpha(L)$.

3.4 Expectiles as an elicitable alternative to Expected Shortfall

Since Value-at-Risk is not coherent and Expected Shortfall lacks direct elicitability, it is interesting to look for risk measures which are coherent as well as elicitable. Possible candidates are Expectiles which we defined in Definition 2.11.

LEMMA 3.1 ([4]) *$e_\tau(L)$ is the unique solution ℓ of the equation*

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

Consequently, $e_\tau(L)$ satifies

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

PROPOSITION 3.4 ([4]) *Expectiles have the following properties:*

(i) For $0 < \tau < 1$, Expectiles are homogeneous and law-invariant.

(ii) For $1/2 \leq \tau < 1$, Expectiles are subadditive (and hence coherent), whereas, for $1/2 \geq \tau > 0$, they are superadditive.

(iii) Expectiles are elicitable.

(iv) Expectiles are additive for linearly dependent random variables, i.e.

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

Bellini and Bignozzi [5] have recently shown that with a slightly narrower definition of elicitability, Expectiles are indeed the only law-invariant and coherent elicitable risk measures.

From Lemma 3.1 and Proposition 3.4, it looks as if Expectiles were ideal to make good for the deficiencies of VaR and ES. This is not the case, however, because Expectiles are not comonotonically additive.

PROPOSITION 3.5 *For $1/2 < \tau < 1$ Expectiles are not comonotonically additive.*

PROOF OF PROPOSITION 3.5.
If e_τ were comonotonically additive then by Theorem 3.6 of [45] it would be a so-called *spectral risk measure*. But then by Corollary 4.3 of [48] it would not be elicitable, in contradiction to Proposition 3.4 (iii). □

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 sub-portfolios or assets the portfolio comprises of. There are quite a few approaches to this problem. See [46] 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 in ([44]) 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 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 $L, L_1, \ldots, L_m$ be random variables such that $L = \sum_{i=1}^m L_i$ and let ρ be a risk measure. If the derivative $\frac{d\rho(L+h\,L_i)}{d\,h}$ exists for $h = 0$ then the risk contribution of L_i to $\rho(L)$ is defined by*

$$\rho(L_i \,|\, L) \;=\; \frac{d\rho(L + h\,L_i)}{d\,h}\bigg|_{h=0}. \tag{4.1}$$

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

$$\rho(L) = \sum_{i=1}^{m} \rho(L_i \,|\, L). \tag{4.2}$$

Tasche in [44] shows that if one of the L_i has a smooth density conditional on the realizations of the other L_i's then the risk contributions of Expected Shortfall 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 in [17] advised an elegant method to determine the risk contributions. In the following theorem we describe the risk contributions to Expected Shortfall. In Theorem 4.2 we then use Delbaen's method to derive the risk contributions to Expectiles.

THEOREM 4.1 ([44], [17]) *If the partial derivative as described in (4.1) exists for ρ chosen as Expected Shortfall, then the risk contribution of a position L_i to the portfolio's Expected Shortfall can be calculated as*

$$\mathrm{ES}_\alpha(L_i|L) = \mathrm{E}[L_i|L \geq q_\alpha(L)] \tag{4.3}$$

With Delbaen's approach, we can also derive the capital allocation for Expectiles.

THEOREM 4.2 *If the partial derivative as described in (4.1) exists for $\rho = e_\tau$, then, for $1/2 \leq \tau < 1$, the risk contribution of a position L_i to the portfolio's Expectile can be calculated as*

$$e_\tau(L_i|L) = \frac{\tau \mathrm{E}[L_i \mathbf{1}_{\{L > e_\tau(L)\}}] + (1-\tau)\mathrm{E}[L_i \mathbf{1}_{\{L \leq e_\tau(L)\}}]}{\tau P[L > e_\tau(L)] + (1-\tau)P[L \leq e_\tau(L)]}. \tag{4.4}$$

PROOF OF THEOREM 4.2.
The sketch of the proof follows Delbaen's method ([17]). Recall that the weak subgradient of a convex function $f : L^\infty(\Omega) \to \mathbb{R}$ at $X \in L^\infty(\Omega)$ (see Section 8.1 of [17]), is defined as:

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

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 [31], e_τ has the so-called Fatou-property (a continuity property),
- as shown in [4], we have that

$$\begin{aligned} e_\tau(L) &= \max\{\mathrm{E}[\varphi\, L] : \varphi \in M_\tau\}, \text{ with} \\ M_\tau &= \Big\{\varphi \geq 0 \text{ is bounded with } \mathrm{E}[\varphi] = 1 \text{ and } \tfrac{\sup \varphi}{\inf \varphi} \leq \max\big(\tfrac{\tau}{1-\tau}, \tfrac{1-\tau}{\tau}\big)\Big\}, \end{aligned}$$

- as shown in [4], for $\bar{\varphi} = \dfrac{\tau\, \mathbf{1}_{\{L > e_\tau(L)\}} + (1-\tau)\, \mathbf{1}_{\{L \leq e_\tau(L)\}}}{\tau\, \mathrm{P}[L > e_\tau(L)] + (1-\tau)\, \mathrm{P}[L \leq e_\tau(L)]}$, we have

$$\bar{\varphi} \in M_\tau \qquad \text{and} \qquad e_\tau(L) = \mathrm{E}[\bar{\varphi}\, L].$$

Theorem 17 of [17] now implies that $\bar{\varphi}$ is an element of $\nabla e_\tau(L)$, i.e. it holds for all bounded random variables L^* that

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

From proposition 5 of [17] it follows that, if $\nabla e_\tau(L)$ has only one element, then we have

$$\left.\frac{d\, e_\tau(L + h\, L^*)}{d\, h}\right|_{h=0} \ = \ \mathrm{E}[\bar{\varphi}\, L^*]. \tag{4.5}$$

Taking $L^* = L_i$ in equation (4.5) implies (4.4). □

REMARK 4.1 *The proof of Theorem 4.2 shows that risk contributions for Expectiles (and also for Expected Shortfall) can still 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 (e.g. 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 [35] 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, indices have been defined, such as the closely related notions of *diversification benefit* defined by Bürgi et al. in [7], and the *diversification index* by Tasche in [46]. Both indices are not universal risk measures and depend on the choice of the risk measure and on the number of the underlying risks in the portfolio.

As mentioned earlier, 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 ([46]):

DEFINITION 4.2 *Let $L_1, \ldots, L_n$ be real-valued random variables and let $L = \sum_{i=1}^{n} L_i$. If ρ is a risk measure such that $\rho(L), \rho(L_1), \ldots, \rho(L_n)$ are defined, then*

$$DI_\rho(L) = \frac{\rho(L)}{\sum_{i=1}^{n} \rho(L_i)} \tag{4.6}$$

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

$$DI_\rho(L_i|L) = \frac{\rho(L_i|L)}{\rho(L_i)} \tag{4.7}$$

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 derived the following properties of the diversification index:

PROPERTIES 4.1 ([46]) *Let ρ be a homogeneous, subadditive, and comonotonically additive risk measure. Then*

- $DI_\rho(L) \leq 1$ *(due to subadditivity).*
- $DI_\rho(L) \approx 1$ *indicates that $L_1, \ldots, L_n$ are 'almost' comonotonic. The closer to one the index of diversification is, the less diversified is the portfolio.*
- *If $DI_\rho(L_i|L) < DI_\rho(L)$, then there exists $\epsilon_i > 0$ such that $DI_\rho(L + hL_i) < DI_\rho(L)$, for all $0 < h < \epsilon_i$.*

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 optimised portfolio might still have a relatively high index. Nonetheless, comparison between marginal diversification indices 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 like 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 lack of diversification due to non-linear dependence.

A notion similar to the diversification index was proposed in [7] 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

$$DB(L) = 1 - \frac{RAC_\rho(\sum_{i=1}^{n} L_i)}{\sum_{i=1}^{n} RAC_\rho(L_i)} \tag{4.8}$$

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:

PROPERTIES 4.2 ([7]) *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*

$$DB(L) = \begin{cases} 1 & \text{indicates full hedging} \\ 0 & \text{indicates comonotonic risks} \\ x \in]0,1[& \text{indicates that there is } 100\,x\% \text{ of capital reduction due to diversification.} \end{cases} \tag{4.9}$$

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 indices depend not only on the choice of ρ and on the portfolio size n, but 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 [8], where it is demonstrated that introducing dependence between the risks drastically reduces the diversification benefits.

When it comes to comparing the consequences of choosing VaR and 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 demonstrated in Proposition 3.3 of [27]. In [8], 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 completely be diversified away.

Moreover, the type of dependence does matter. Linear dependence (measured with the linear correlation) cannot accurately describe dependence between extreme risks, in particular in times of stress. Neglecting the non-linearity of dependence may lead to an overestimation of the diversification benefits. This is well described by Bürgi et al. [7] who consider elliptical and Archimedean copulae for risk modelling and compare their impacts on the evaluation of RAC and hence also on the diversification benefit.

5 Backtesting: which methods can be used?

What does backtesting mean? According to Jorion [30], it is a set of statistical procedures designed to check if the real losses, observed ex post, are in line with VaR forecasts. We may of course extend this definition to any risk measure.

Recently, Gneiting ([28]) has raised the issue of direct backtesting when using Expected Shortfall (ES) as a risk measure. This is not an issue for risk measures like VaR or Expectiles because of their elicitability, as seen previously. Is it a real issue in practice for ES? Some financial institutions seem to have circumvented this problem even when using the risk measure ES. In the following, we discuss which backtest methods can be used in practice. In particular, we propose an empirical approach that consists in approximating ES with quantiles – which allows to make use of backtesting methods for VaR.

As observed in Section 3.3, ES is a combination of two elicitable components. A natural approach to the backtesting of ES therefore is to backtest both components separately according to their associated respective scoring functions (described in [28]).

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

(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 (e.g. [12] or [21]).

(ii) *Probability range forecasts* or *interval forecasts* (e.g. forecasts of Value-at-Risk or of Expected Shortfall); they project an interval in which the forecast value is expected to lie with some probability p (e.g. the interval $(-\infty, VaR_p(Y_{t+k})]$ where $VaR_p(Y_{t+k})$ 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 15 years. A good reference on this topic is Christoffersen ([11]). Backtesting for VaR has been well developed, due to the interest of the financial industry in this risk measure. We refer e.g. to Davé et al. ([16]), and, for a review on backtesting procedures for the VaR, to Campbell ([9]).

(iii) *Forecasts of the complete probability distribution* $\mathbb{P}[Y_{t+k} \leq . \mid \mathcal{F}(Y_s, s \leq t)]$ or its probability density function, if existing.

It is worth noticing that if there is a solution to (iii) then there are also solutions for (i) and (ii), and that (iii) makes it possible to backtest ES, avoiding then the issue raised by Gneiting ([28]) 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. 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 [38] or [22]).

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

5.1 Backtesting VaR and ES

Backtesting VaR. 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 $\mathbb{P}(L > VaR_\alpha(L)) = 1-\alpha$, it follows that the probability of a violation of VaR is $1-\alpha$. We define the *violation process* of VaR as

$$I_t(\alpha) = \mathbf{1}_{\left\{L(t) > VaR_\alpha(L(t))\right\}}. \tag{5.1}$$

Here $\mathbf{1}$ denotes the indicator function of the event $\{L(t) > VaR_\alpha(L(t))\}$.

Christoffersen ([11]) showed that VaR forecasts are valid if and only if the violation process $I_t(\alpha)$ satisfies two conditions:

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

Under these two conditions, the $I_t(\alpha)$'s are independent and identically distributed Bernoulli random variables with success probability $1-\alpha$. Hence the number of violations has a Binomial distribution.

In practice, we want to compare VaR predictions with observed values to assess the quality of the predictions. To do so, we consider an estimate of the violation process by replacing VaR by its estimates and check that this process behaves like independent and identically distributed Bernoulli random variables with violation (success) probability close to $1-\alpha$. If the proportion of VaR violations is not significantly different from $1-\alpha$, then we conclude that the estimation/prediction method is reasonable.

Backtesting ES. A similarly simple approximative approach to the backtesting of ES can be based on a representation of ES as integrated VaR ([1], Proposition 3.2):

$$\begin{aligned} \mathrm{ES}_\alpha(L) &= \frac{1}{1-\alpha}\int_\alpha^1 q_u(L)\,du \\ &\approx \frac{1}{4}\left[\, q_\alpha(L) + q_{0.75\,\alpha+0.25}(L) + q_{0.5\,\alpha+0.5}(L) + q_{0.25\,\alpha+0.75}(L)\,\right], \end{aligned} \tag{5.2}$$

where $q_\alpha(L) = VaR_\alpha(L)$. 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 to some extent also the estimate of $\mathrm{ES}_\alpha(L)$ can be considered reliable.
This approach is attractive not only for 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 $\mathrm{ES}_\alpha(L)$ than for $\mathrm{VaR}_\alpha(L)$ (see also [47]). The Basel Committee suggests a variant of this ES-backtesting approach which is based on testing level violations for two quantiles at 97.5% and 99% level [3].

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, named also *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 like ES.

The use of the PIT for backtesting financial models is recent. The foundations were laid by Diebold and coauthors. Diebold et al. in [19] tackled the problem of density forecast evaluation from a risk management perspective, suggesting a method for testing distribution forecasts in finance, based on the uniform distribution of the Lévy-Rosenblatt transform (or PIT) ([34] and [41]). Applying the Lévy theorem to the PIT, they observed that if a sequence of distribution forecasts coincides with the sequence of unknown conditional laws that have generated the observations, then the sequence of PIT are independent and identically distributed $\mathcal{U}(0,1)$. In [20], 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.

Nevertheless, there was still some gap to fill up before a full implementation and use in practice. Blum in his PhD thesis ([6]) studied various issues left open, and proposed and validated mathematically a method based on PIT also in situations with overlapping forecast intervals and multiple forecast horizons. Blum illustrated this in his thesis dealing with economic scenario generators (ESG). Typically, financial institutions make use of scenario generators, producing thousands of scenarios, each one having its own forecast value for a certain value at a certain future time. Those values define an empirical distribution, which represents a distribution forecast. Hence the backtesting will be done on the modeling distribution; it is an out-of-sample backtesting of distribution forecasts. For details of the methodology, we refer to [6], [42] and the references therein and only summarize the main steps in the following.

The asymptotic limit distribution of many scenarios converges to the marginal cdf $\Phi_i(x) = \mathbb{P}(X_i < x \mid \mathcal{F}_{i-m})$ where X_i corresponds to the scenario forecast of a variable X at out-of-sample time point t_i and $\mathcal{F}_{i-m}$ to the information available up to time t_{i-m} from the simulation start, m being the number of forecast steps.

The empirical distribution $\hat{\Phi}_i$ can be shown to be close to Φ, when taking many scenarios. Hence, identifying them, we have at t_i a distribution forecast $\hat{\Phi}_i(x)$ and a historically observed value x_i.

Now we apply the PIT to build the random variables $Z_i := \hat{\Phi}_i(x_i)$. These have been proved by Diebold et al. ([19], [20]) to be independent and identically $\mathcal{U}(0,1)$-distributed whenever the conditional distribution forecast $\Phi_i(.)$ coincides with the true process by which the historical data have 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 ([18]), 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. Going on with the Diebold et al. methodology, their non-parametric test, proposed in [19] (see also [20] 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 of detecting deviations from the independence property when considering correlograms of the Z_i and their lower integer powers.

Note that tests based on PIT have some limitation due to serial correlation. One way to overcome this issue is for instance, as suggested in [42], to generate realistic forecast scenarios via refined bootstrapping.

6 Conclusion

In this paper, we have listed a number of properties that are commonly considered must-haves for good risk measures: coherence, comonotonic additivity, robustness, and elicitability. We have then revisited the popular risk measures Value-at-Risk (VaR) and Expected Shortfall (ES) as well as the recently suggested Expectiles and checked which of these 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 well known to be sub-optimal in terms of risk for VaR (e.g. [37], Example 6.7).

- ES makes good for the lack of subadditivity and sensitivity for tail risk of VaR but has recently be found to be not elicitable. This means that backtesting of ES is less straightforward than backtesting of VaR. We have found that nonetheless there are a number of feasible approaches to the backtesting of ES although 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 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 non-linear dependencies.

To conclude, we have found that among the risk measures we discussed, ES seems 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 in mind Expectiles as alternatives to ES and VaR in specific applications.

References

[1] C. Acerbi, D. Tasche, On the coherence of expected shortfall. *Journal of Banking & Finance* **26**, (2002) 1487-1503.

[2] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risks. *Mathematical Finance* **9**, (1999) 203-228.

[3] BCBS. *Fundamental review of the trading book: A revised market risk framework.* Basel Committee on Banking Supervision, October 2013.

[4] F. Bellini, B. Klar, A. Müller, and E. Rosazza Gianin, Generalized quantiles as risk measures. Preprint (2013) URL http://ssrn.com/abstract=2225751.

[5] F. Bellini, V. Bignozzi, Elicitable Risk Measures (2013). Available at SSRN: http://ssrn.com/abstract=2334746

[6] P. Blum, On some mathematical aspects of dynamic financial analysis. PhD thesis, ETH Zürich (Swiss Federal Institute of Technology, Zurich) (2004).

[7] R. Bürgi, M.M. Dacorogna, and R. Iles, Risk Aggregation, dependence structure and diversification benefit. *In: Stress testing for financial institutions, Riskbooks, Incisive Media, London*, (2008) 265-306.

[8] M. Busse, M.M. Dacorogna, and M. Kratz, The impact of systemic risk on the diversification benefits of a risk portfolio. *ESSEC working paper* **1321**, (2013) 1-17, & *preprint* arXiv:1312.0506.

[9] S.D. Campbell, A review of Backtesting and Backtesting Procedures. *Journal of Risk* **9(2)**, (2006) 1-17.

[10] J.M. Chen, Measuring Market Risk Under Basel II, 2.5, and III: VAR, Stressed VAR, and Expected Shortfall (2013). Available at SSRN: http://ssrn.com/abstract=2252463

[11] P. Christoffersen, *Elements of Financial Risk Management.* Academic Press (2003).

[12] M. Clements, D. Hendry, *Forecasting Economic Time Series*. Cambridge University Press (1998).

[13] R. Cont, R. Deguest, and G. Scandolo, Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance* **10(6)**, (2010) 593-606.

[14] H. Cramér, *On the Mathematical Theory of Risk*. Skandia Jubilee Volume, Stockholm (1930).

[15] J. Daníelson, B. Jorgenson, G. Samorodnitsky, M. Sarma, C. de Vries, Fat tails, VaR and subadditivity. *Journal of Econometrics* **172(2)**, (2013) 283-291.

[16] R.D. Davé, G. Stahl, On the Accuracy of VaR Estimates Based on the Variance-Covariance Approach. *Risk Measurement, Econometrics and Neural Networks*, (1998), 189-232.

[17] F. Delbaen, Coherent risk measures. PISA Lecture notes (2000)
URL http://www.math.ethz.ch/∼delbaen/ftp/preprints/PISA007.pdf

[18] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**, (1995) 253-263.

[19] F.X. Diebold, T.A. Gunther, and A.S. Tay, Evaluating density forecasts with application to financial risk management. *Int. Econ. Rev.* **39**, (1998) 863-883.

[20] F.X. Diebold, J. Hahn, and A.S. Tay Multivariate density forecasts and calibration in financial risk management: High-frequency returns on foreign exchange. *The Review of Economics and Statistics* **81**, (1999) 661-673.

[21] G. Elliott, C. Granger and A. Timmermann, *Handbook of Economic Forecasting*, Vol. 1 (2006) & Vol. 2 Elsevier (2013)

[22] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*. Springer (1997).

[23] P. Embrechts, D. Lambrigger, and M. Wüthrich, Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes* **12** (2009).

[24] P. Embrechts, A. McNeil, and D. Straumann, Correlation and dependence in risk management: properties and pitfalls. In *M. Dempster and A. Howarth, editors, Risk management: value at risk and beyond*, Cambridge University Press, (2002) 176-223.

[25] P. Embrechts, J. Neslehová, and M. Wüthrich, Additivity properties for Value-at-Risk under Archimedean dependence and heavy-tailedness. *Insurance: Mathematics and Economics* **44**, (2009) 164-169

[26] P. Embrechts, G. Puccetti, and L. Rüschendorf, Model uncertainty and VaR aggregation. *Journal of Banking & Finance* **37(8)**, (2013) 2750-2764.

[27] S. Emmer, D. Tasche, Calculating Credit Risk Capital Charges with the One-factor Model. *Journal of Risk* **7(2)**, (2004) 85-103.

[28] T. Gneiting, Making and evaluating point forecasts. *Journal of the American Statistical Association* **106(494)** (2012), 746-762.

[29] A. Hauksson, M.M. Dacorogna, T. Domenig, U. A. Müller and G. Samorodnitsky, Multivariate Extremes, Aggregation and Risk Estimation. *Quantitative Finance* **1** (2001) 79-95.

[30] P. Jorion, *Financial Risk Manager Handbook*, Wiley, 4th ed. (2007).

[31] E. Jouini, W. Schachermayer and N. Touzi, Law invariant risk measures have the Fatou property. *Advances in mathematical economics*, Springer Japan (2006) 49-71.

[32] M. Kratz. There is a VaR beyond usual approximations. Towards a toolkit to compute risk measures of aggregated heavy tailed risks. *FINMA report* (2013).

[33] M. Kratz, There is a VaR beyond usual approximations. *ESSEC working paper* **1317**, (2013) 1-33, & *preprint* arXiv:1311.0270.

[34] P. Lévy, *Théorie de l'addition des variables aléatoires*. Gauthier-Villars, Paris, (1937).

[35] M. Kalkbrener, An axiomatic approach to capital allocation. *Mathematical Finance* **15(3)**, (2005) 425-437.

[36] N.S. Lambert, D.M. Pennock and Y. Shoham, Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, ACM (2008) 129-138.

[37] A. McNeil, R. Frey, P. Embrechts, *Quantitative Risk Management*. Princeton (2005).

[38] J. Pickands, Statistical inference using extreme order statistics. *Ann. Stat.* **3**, (1975) 119-131.

[39] H. Markowitz, Portfolio selection, *The Journal of Finance* **7(1)**, (1952).

[40] R.T. Rockafellar and S. Uryasev, The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science* **18(1)**, (2013), 33-53.

[41] M. Rosenblatt, Remarks on a multivariate transformation. *The Annals of Mathematical Statistics* **23**, (1952) 470-472.

[42] SCOR Switzerland, From Principle-Based Risk Management to Solvency Requirements. Analytical Framework for the Swiss Solvency Test (2008).

[43] G. Stahl, J. Zheng, R. Kiesel, R. Rühlicke, Conceptualizing Robustness in Risk Management. 2012 SSRN: http://ssrn.com/abstract=2065723.

[44] D. Tasche, Risk contributions and performance measurement. (1999) Working paper, Technische Universität München.

[45] D. Tasche, Expected Shortfall and Beyond. *Journal of Banking & Finance* **26(7)**, (2002) 1519-1533.

[46] D. Tasche, Capital allocation to business units and sub-portfolios: the Euler principle. In: Resti, A., editor, *Pillar II in the New Basel Accord: The Challenge of Economic Capital*, Risk Books, (2008) 423-453.

[47] Y. Yamai, T. Yoshiba Value-at-Risk versus Expected Shortfall: A practical perspective. *Journal of Banking & Finance* **29(4)**, (2005) 997-1015.

[48] J.F. Ziegel, Coherence and Elicitability. (2013) *preprint* arXiv:1303.1690.

ESSEC Business School
Avenue Bernard Hirsch
BP 50105
95021 Cergy-Pontoise Cedex
France

Tél. +33 (0)1 34 43 30 00
Fax +33 (0)1 34 43 30 01

www.essec.fr

ESSEC Executive Education
CNIT BP 230
92053 Paris-La Défense
France

Tél. +33 (0)1 46 92 49 00
Fax +33 (0)1 46 92 49 90

http://formation.essec.fr

ESSEC Business School
Singapore Campus
100 Victoria Street
National Library Building # 13-02
Singapore 188064

essecasia@essec.fr
Tél. +65 6884 9780
Fax +65 6884 9781

www.essec.edu

Informations

Alison Bougi
+33 (0)1 34 43 33 58
bougi@essec.fr
www.essec.fr
research.center@essec.fr

ISSN 1291-9616

ESSEC BUSINESS SCHOOL

CCI À vos côtés, précisément. AACSB EFMD EQUIS ACCREDITED