Elicitability of Range Value at Risk

Tobias Fissler* Johanna F. Ziegel†

March 27, 2019

Abstract. The predictive performance of point forecasts for a statistical functional, such as the mean, a quantile, or a certain risk measure, is commonly assessed in terms of scoring (or loss) functions. A scoring function should be (strictly) consistent for the functional of interest, that is, the expected score should be minimised by the correctly specified functional value. A functional is elicitable if it possesses a strictly consistent scoring function. In quantitative risk management, the elicitability of a risk measure is closely related to comparative backtesting procedures. As such, it has gained considerable interest in the debate about which risk measure to choose in practice. While this discussion has mainly focused on the dichotomy between Value at Risk (VaR)—a quantile—and Expected Shortfall (ES)—a tail expectation, this paper is concerned with Range Value at Risk (RVaR). RVaR can be regarded as an interpolation of VaR and ES, which constitutes a tradeoff between the sensitivity of the latter and the robustness of the former. Recalling that RVaR is not elicitable, we show that a triplet of RVaR with two VaR components at different levels is elicitable. We characterise the class of strictly consistent scoring functions. Moreover, additional properties of these scoring functions are examined, including the diagnostic tool of Murphy diagrams. The results are illustrated with a simulation study, and we put our approach in perspective with respect to the classical approach of trimmed least squares in robust regression.

Keywords: Consistency; Interquantile expectation; Point forecasts; Robustness; Scoring functions; Trimmed mean; Value at Risk; Winsorized mean

AMS 2010 Subject Classification: 62G35; 62P05; 91G70

*Imperial College London, Department of Mathematics, Section Statistics, 180 Queen's Gate, London SW7 2AZ, United Kingdom, e-mail: `t.fissler@imperial.ac.uk`

†University of Bern, Department of Mathematics and Statistics, Institute of Mathematical Statistics and Actuarial Science, Alpeneggstrasse 22, 3012 Bern, Switzerland, e-mail: `johanna.ziegel@stat.unibe.ch`

1 Introduction

1.1 Mean vs. median in classical statistics

The two most prominent measures for the location of a distribution are the mean and the median. Both of them have a clear and accessible interpretation. While they coincide for symmetric distributions, they can considerably differ for asymmetric ones. From an estimation point of view the difference between the two measures are even more pronounced: The population mean is sensitive with respect to the underlying distribution, and—for symmetric distributions—it is a more efficient location-estimator than the median for light-tailed distributions (Koenker and Basset, 1978). On the other hand, the median is esteemed for its robustness against outliers, and—again for symmetric distributions—it turns out to be a more efficient location-estimator than the mean for heavy-tailed distributions (*ibidem*). Indeed, one can show that the maximum likelihood estimator for location coincides with the sample mean if the underlying distribution is normal whereas it amounts to the sample median in case of a Laplace distribution (Keynes, 1911). Moreover, the median of a distribution always exists while the existence of the mean requires a benign tail behaviour of the distribution.

The field of *robust statistics* was started off by the seminal contributions of Tukey (1960) and Huber (1964). Hampel (1971) was the first to formalise the notion of robustness and to link it to a continuity property of the estimator. Besides this qualitative definition of robustness, Hampel also introduced the *breakdown point* of an estimator as a quantitative measure of robustness. In finite samples (Donoho and Huber, 1983), it roughly amounts to the proportion of data that can be changed without corrupting the estimator. As a consequence, the median is robust with a breakdown point of 1/2, whereas the mean achieves a breakdown point of 0 rendering it non-robust. Since the early days of robust statistics, the field has developed an incredibly rich strand of literature. For a thorough introduction we refer the reader to the excellent textbook Huber and Ronchetti (2009).

It is well known that both the mean and the median can be expressed as M-estimators using the squared loss or the absolute loss, respectively. The most prominent compromise between the mean and the median in form of an M-estimator is given by the famous Huber loss (Huber, 1964, p. 79). Historically older alternatives of such a compromise are the α-*trimmed mean* and the α-*Winsorized mean*, belonging to the class of L-estimates. On a population level, the α-trimmed mean, $\alpha \in (0, 1/2)$, is the average of all β-quantiles for $\beta \in [\alpha, 1-\alpha]$ (see Section 2 for precise definitions). In a finite sample, it amounts to removing the smallest and the largest α-fraction of all observations and then computing the mean with the remaining $(1-2\alpha)$-fraction of observations. The α-Winsorized mean instead calculates the mean over all observations, with the smallest (largest) α-fraction set to be the empirical α-quantile ($(1-\alpha)$-quantile). As such, the α-trimmed mean and the α-Winsorized mean constitute two natural interpolations between the mean ($\alpha = 0$) and the median ($\alpha = 1/2$). They are robust, with a breakdown point of α (Hampel, 1971).

1.2 Expected Shortfall vs. Value at Risk in risk management

In the field of quantitative risk management, the last one or two decades have seen a lively debate about which monetary risk measure (Artzner et al., 1999) be best in (regulatory) practice. The debate mainly focused on the dichotomy between Value at Risk (VaR_α) on the one hand and Expected Shortfall (ES_α) on the other hand at some level $\alpha \in (0,1)$ (see Section 2 for definitions). Interestingly, and in line with the debate in classical statistics, this encompasses a joust between a quantile (VaR_α) and a tail expectation (ES_α). We refer the reader to Embrechts et al. (2014) and Emmer et al. (2015) for comprehensive academic discussions and to Bank for International Settlements (2014) for a regulatory perspective in banking.

Cont et al. (2010) considered the issue of statistical robustness of risk measure estimates in the sense of Hampel (1971). They showed that a risk measure cannot be both robust in the latter sense and *coherent* in the sense of Artzner et al. (1999). As a compromise, they propose the risk measure 'Range Value at Risk' ($\mathrm{RVaR}_{\alpha,\beta}$), which is akin to an asymmetric version of the trimmed mean: One takes the average of all quantiles between two extreme levels $0 \le \alpha \le \beta \le 1$, rather than between two symmetric levels where $\beta = 1-\alpha$ (see Section 2 for definitions). Setting $\alpha = \beta \in (0,1)$ renders VaR_β and $\alpha = 0 < \beta < 1$ leads to ES_β. The arguments provided in Huber and Ronchetti (2009, p. 59) imply that $\mathrm{RVaR}_{\alpha,\beta}$ has a breakdown point of $\min\{\alpha, 1-\beta\}$, which means it is a robust—and hence, not coherent—risk measure, unless it degenerates to $\mathrm{RVaR}_{0,\beta} = \mathrm{ES}_\beta$ (or if $0 \le \alpha < \beta = 1$). Moreover, RVaR belongs to the wide class of distortion risk measures (Kusuoka, 2001). For further contributions to robustness in the context of risk measures, we refer the reader to Krätschmer et al. (2012, 2014), Kou et al. (2013), Embrechts et al. (2015) and Zähle (2016). Since the influential article Cont et al. (2010), RVaR has gained increasing attention in the risk management literature—see Embrechts et al. (2018a,b) for extensive studies—as well as in econometrics (Barendse, 2017) where RVaR sometimes has the alternative denomination *Interquantile Expectation.*

1.3 Elicitability

The property of a statistical functional to have an M-estimator on the population level has become known as *elicitability* (Osband, 1985; Lambert et al., 2008; Gneiting, 2011). More specifically, we say that a functional T is elicitable if there is a scoring function $S(x,y)$ such that $T(F) = \arg\min_x \int S(x,y)\,\mathrm{d}F(y)$. *Vice versa,* a scoring function is called *strictly consistent* for T if its expectation is uniquely minimised in x at $T(F)$. Examples for elicitable functionals are given by the mean with $S(x,y) = (x-y)^2$ and the median with $S(x,y) = |x-y|$ and their asymmetric versions, expectiles and quantiles. From a game theoretic point of view, strict consistency of a scoring function amounts to incentive compatibility, rewarding truthful and honest forecasts. Besides its importance for M-estimation and regression, e.g. quantile regression (Koenker and Basset, 1978; Koenker, 2005) or expectile regression (Newey and Powell, 1987), the notions of elicitability and strict consistency are crucial for forecast evaluation (Engelberg et al., 2009; Murphy and Daan, 1985). If forecasts take the form of probability distributions or densities, one often uses the term scoring rule rather than scoring function and propriety rather than consistency (Gneiting and Raftery, 2007).

Osband (1985) showed that convex level sets (CxLS) of a functional are necessary for its elicitability. This shows that variance is generally not elicitable, and also Expected Shortfall fails to have the CxLS-property (Weber, 2006; Gneiting, 2011). Steinwart et al. (2014) showed that for continuous one-dimensional functionals, the CxLS-property is basically also sufficient for elicitability; cf. Lambert (2013), Bellini and Bignozzi (2015), Delbaen et al. (2016), as well as Heinrich (2014) for the role of the continuity assumption. The *revelation principle* (Osband, 1985; Gneiting, 2011) asserts that any bijection of an elicitable functional is elicitable. This implies that the pair (mean, variance)—being a bijection of the first two moments—is elicitable despite the variance fails to be elicitable. Similarly, Fissler and Ziegel (2016) showed that the pair $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ is elicitable with the structural difference that the revelation principle is not applicable in this instance. This gave rise to the finding that the minimal expected score and its minimiser are jointly elicitable; see Frongillo and Kash (2015) and Brehmer (2017).

In the context of quantitative finance and particularly in the debate about which risk measure is best in practice, elicitability has gained considerable attention (Emmer et al., 2015; Ziegel, 2016; Davis, 2016). Especially, the role of elicitability for backtesting purposes has been highly debated (Gneiting, 2011; Acerbi and Székely, 2014, 2017; Fissler et al., 2016; Nolde and Ziegel, 2017).

1.4 Elicitability of Range Value at Risk

Very recently, Wang and Wei (2018) showed that $\mathrm{RVaR}_{\alpha,\beta}$, $0 < \alpha < \beta < 1$, similarly to ES_α, fails to have the CxLS property, which rules out its elicitability (see Proposition 3.1). In contrast, they observe that the identity

$$\mathrm{RVaR}_{\alpha,\beta} = \big(\beta\,\mathrm{ES}_\beta - \alpha\,\mathrm{ES}_\alpha\big)/(\beta-\alpha), \qquad 0<\alpha<\beta<1, \tag{1.1}$$

and the CxLS property of the pair $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ implies the CxLS property of the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ (Wang and Wei, 2018, Example 7), leading to the question whether this triplet is elicitable or not. Invoking the elicitability of $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$, the identity at (1.1) and the revelation principle establishes the elicitability of the quadruples $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{ES}_\alpha, \mathrm{RVaR}_{\alpha,\beta})$ and $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{ES}_\beta, \mathrm{RVaR}_{\alpha,\beta})$. This approach has already been used in the context of regression in Barendse (2017).

A fortiori, we show that the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ is elicitable (Theorem 3.4) under weak regularity conditions. Besides the obvious advantage that this reduces the elicitation complexity (Lambert et al., 2008; Frongillo and Kash, 2015) or elicitation order (Fissler and Ziegel, 2016), it is particularly superior since $\mathrm{RVaR}_{\alpha,\beta}(F)$, $0 < \alpha < \beta < 1$, exists for any distribution F, while $\mathrm{ES}_\alpha(F)$ and $\mathrm{ES}_\beta(F)$ only exist if the (left) tail of the distribution F is integrable. Since RVaR is used often for robustness purposes, safeguarding against outliers and heavy-tailedness, the latter advantage becomes particularly important.

We would like to point out the structural difference between the elicitability result of $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ provided in this paper and the one concerning $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ in Fissler and Ziegel (2016) as well as the more general results of Frongillo and Kash (2015) and Brehmer (2017). While ES_α corresponds to the negative of a minimum of an expected score which is strictly consistent for VaR_α, it turns out that $\mathrm{RVaR}_{\alpha,\beta}$ can be represented as the *difference* of minima of strictly consistent scoring functions for VaR_α and VaR_β, respectively (Lemma 3.3). As a consequence, the class of strictly consistent scoring functions for the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ turns out to be less flexible than the one for $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$; see Remark 3.9 for details. One particular implication is that there are essentially no strictly consistent scoring functions for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ which are also translation invariant or positively homogeneous; see Section 4.

The paper is organised as follows. In Section 2, we introduce the relevant notation and definitions concerning RVaR, scoring functions and elicitability. The main results establishing the elicitability of the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ (Theorems 3.4 and 3.7) and related findings are presented in Section 3. Section 4 shows that there are basically no strictly consistent scoring functions for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ which are positively homogeneous or translation invariant. In Section 5, we establish a mixture representation of the strictly consistent scoring functions in the spirit of Ehm et al. (2016). This result allows to compare forecasts simultaneously with respect to *all* consistent scoring functions in terms of *Murphy diagrams*. We demonstrate the applicability of our results and compare the discrimination ability of different scoring functions in a simulation study presented in Section 6. The paper finishes in Section 7 with a discussion of our results in the context of M-estimation and compares them to other suggestions in the statistical literature, in variants of a *trimmed least squares* procedure (Koenker and Basset, 1978; Ruppert and Carroll, 1980; Rousseeuw, 1984). A list of assumptions similar to the ones used in Fissler and Ziegel (2016) can be found in the Appendix.

2 Notation and Definitions

2.1 Definition of Range Value at Risk

We would like to recall that there are different sign conventions in the literature about risk measures. In this paper we use the following convention: If a random variable Y models the losses and gains, then positive values of Y represent gains and negative values of Y losses. Since we consider law-invariant risk measures only, thus defining risk measures directly as functionals of the distribution of Y, corresponding comments apply. Moreover, if ρ is a risk measure, we assume that $\rho(Y) \in \mathbb{R}$ corresponds to the maximal amount of money one can *withdraw* such that the position $Y - \rho(Y)$ is still acceptable. Hence, negative values of ρ correspond to risky positions.

Definition 2.1 (Value at Risk)**.** Let F be a probability distribution function on $\mathbb{R}$. For any $\alpha \in (0,1)$ we define the *Value at Risk* of F at level α via

$$\mathrm{VaR}_\alpha(F) = \inf\{x \in \mathbb{R} \,|\, \alpha \le F(x)\} \in \mathbb{R}\,.$$

Moreover, we use the common convention that $\mathrm{VaR}_0(F) \in \mathbb{R} \cup \{-\infty\}$ corresponds to the infimum of the support of F and that $\mathrm{VaR}_1(F) \in \mathbb{R} \cup \{\infty\}$ is defined as the supremum of the support of F.

Definition 2.2 (Range Value at Risk)**.** Let F be a probability distribution function on $\mathbb{R}$. For $0 \le \alpha \le \beta \le 1$ we define the *Range Value at Risk* of F at levels α, β via

$$\mathrm{RVaR}_{\alpha,\beta}(F) = \begin{cases} \dfrac{1}{\beta-\alpha}\displaystyle\int_\alpha^\beta \mathrm{VaR}_\gamma(F)\,\mathrm{d}\gamma, & \text{if } \alpha < \beta, \\ \mathrm{VaR}_\alpha(F), & \text{if } \alpha = \beta. \end{cases}$$

Note that our parametrisation of $\mathrm{RVaR}_{\alpha,\beta}$ differs from the one in Embrechts et al. (2018b). One can verify that $\mathrm{RVaR}_{\alpha,\beta}(F) \in \mathbb{R}$ if $0 < \alpha \le \beta < 1$. For $\beta \in (0,1)$, $\mathrm{RVaR}_{0,\beta}(F) \in \mathbb{R} \cup \{-\infty\}$ and it is finite if and only if $\int_{-\infty}^0 |y|\,\mathrm{d}F(y) < \infty$. Similarly, for $\alpha \in (0,1)$ it holds that $\mathrm{RVaR}_{\alpha,1}(F) \in \mathbb{R} \cup \{\infty\}$ and it is finite if and only if $\int_0^\infty |y|\,\mathrm{d}F(y) < \infty$. $\mathrm{RVaR}_{0,1}(F)$ exists only if $\int_{-\infty}^0 |y|\,\mathrm{d}F(y) < \infty$ or $\int_0^\infty |y|\,\mathrm{d}F(y) < \infty$. If F has a finite first moment, then $\mathrm{RVaR}_{0,1}(F) = \int y\,\mathrm{d}F(y)$ coincides with the first moment of F.

One can generalise this identity for $0 \le \alpha < \beta \le 1$ and obtains the alternative representation

$$\mathrm{RVaR}_{\alpha,\beta}(F) = \frac{1}{\beta-\alpha}\Bigg(\int_{(\mathrm{VaR}_\alpha(F),\mathrm{VaR}_\beta(F)]} y\,\mathrm{d}F(y) \\ + \mathrm{VaR}_\alpha(F)\big(F(\mathrm{VaR}_\alpha(F)) - \alpha\big) - \mathrm{VaR}_\beta(F)\big(F(\mathrm{VaR}_\beta(F)) - \beta\big)\Bigg),$$

where we used the usual convention that $F(-\infty) = 0$, $F(\infty) = 1$ and $0\cdot\infty = 0\cdot(-\infty) = 0$. If F is continuous at its α- and β-quantiles in the sense that $F(\mathrm{VaR}_\alpha(F)) = \alpha$ and $F(\mathrm{VaR}_\beta(F)) = \beta$ then the correction terms in (2.2) vanish and one has that

$$\mathrm{RVaR}_{\alpha,\beta}(F) = \frac{1}{\beta-\alpha}\int_{(\mathrm{VaR}_\alpha(F),\mathrm{VaR}_\beta(F)]} y\,\mathrm{d}F(y) \\ = \frac{1}{\beta-\alpha}\mathbb{E}_F[Y\,\mathbb{1}\{\mathrm{VaR}_\alpha(F) < Y \le \mathrm{VaR}_\beta(F)\}],$$

which justifies an alternative name for RVaR, namely *Interquantile Expectation*.

Definition 2.3 (Expected Shortfall)**.** Let F be a probability distribution function on $\mathbb{R}$. For any $\alpha \in (0,1)$ we define the *Expected Shortfall* of F at level α via

$$\mathrm{ES}_\alpha(F) = \mathrm{RVaR}_{0,\alpha}(F) \in \mathbb{R} \cup \{-\infty\}.$$

Let $0 < \alpha < \beta < 1$ and $\mathrm{ES}_\alpha(F), \mathrm{ES}_\beta(F) \in \mathbb{R}$. Then one obtains the identity

$$\mathrm{RVaR}_{\alpha,\beta}(F) = \big(\beta\,\mathrm{ES}_\beta(F) - \alpha\,\mathrm{ES}_\alpha(F)\big)/(\beta-\alpha). \tag{2.1}$$

If F has a finite left tail ($\int_{-\infty}^0 |y| \,\mathrm{d}F(y) < \infty$) then one could use the right hand side of (2.1) as a definition of $\mathrm{RVaR}_{\alpha,\beta}(F)$. However, in line with our discussion in the introduction, $\mathrm{RVaR}_{\alpha,\beta}(F)$ always exists and is finite for $0 < \alpha < \beta < 1$ even if the right hand side of (2.1) is not defined.

Interestingly, Theorem 2 in Embrechts et al. (2018b) establishes that RVaR can be written as inf-convolution of VaR and ES at appropriate levels. Note that this would rather amount to a sup-convolution in our context due to different sign conventions.

For $\alpha \in (0, 1/2)$, $\mathrm{RVaR}_{\alpha,1-\alpha}$ corresponds to the α-*trimmed mean* and has a close connection to the α-*Winsorized mean* W_α (Huber and Ronchetti, 2009, pp. 57–59) via

$$W_\alpha(F) := (1-2\alpha)\,\mathrm{RVaR}_{\alpha,1-\alpha}(F) + \alpha\,\mathrm{VaR}_\alpha(F) + \alpha\,\mathrm{VaR}_{1-\alpha}(F), \qquad \alpha \in (0, 1/2). \quad (2.2)$$

It is easy to verify that for any distribution function F and $0 < \alpha < \beta < 1$ one obtains the inequality

$$\mathrm{VaR}_\alpha(F) \le \mathrm{RVaR}_{\alpha,\beta}(F) \le \mathrm{VaR}_\beta(F). \quad (2.3)$$

2.2 Elicitability and scoring functions

We essentially follow the notation used in Fissler and Ziegel (2016), which follows the decision-theoretic framework used in Gneiting (2011). Let $\mathcal{F}$ be a generic class of probability distribution functions on $\mathbb{R}$. An *action domain* A is a subset $\mathsf{A} \subseteq \mathbb{R}^k$ for $k \ge 1$. Whenever we consider a functional $T\colon \mathcal{F} \to \mathsf{A}$, we tacitly assume that $T(F)$ is well-defined for all $F \in \mathcal{F}$ and is an element of A. $T(\mathcal{F})$ corresponds to the image $\{T(F) \in \mathsf{A} \,|\, F \in \mathcal{F}\}$. For any subset $M \subseteq \mathbb{R}^k$ we denote with $\mathrm{int}(M)$ the largest open subset of M. Moreover, $\mathrm{conv}(M)$ denotes the convex hull of the set M.

We say that a function $a\colon \mathbb{R} \to \mathbb{R}$ is $\mathcal{F}$-integrable if it is measurable and $\int |a(y)| \,\mathrm{d}F(y) < \infty$ for all $F \in \mathcal{F}$. Similarly, a function $g\colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ is called $\mathcal{F}$-integrable if $g(x, \cdot)\colon \mathbb{R} \to \mathbb{R}$ is $\mathcal{F}$-integrable for all $x \in \mathsf{A}$. If g is $\mathcal{F}$-integrable, we define the map

$$\bar{g}\colon \mathsf{A} \times \mathcal{F} \to \mathbb{R}, \quad \bar{g}(x, F) := \int g(x, y) \,\mathrm{d}F(y).$$

If $g\colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ is sufficiently smooth in its first argument, we denote the mth partial derivative of $g(\cdot, y)$ with $\partial_m g(\cdot, y)$.

Definition 2.4 (Consistency and elicitability)**.** A *scoring function* is an $\mathcal{F}$-integrable map $S\colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}$. It is called $\mathcal{F}$-*consistent* for a functional $T\colon \mathcal{F} \to \mathsf{A}$ if $\bar{S}(T(F), F) \le \bar{S}(x, F)$ for all $x \in \mathsf{A}$ and for all $F \in \mathcal{F}$. It is *strictly* $\mathcal{F}$-consistent for T if it is consistent and if $\bar{S}(T(F), F) = \bar{S}(x, F)$ implies that $x = T(F)$ for all $x \in \mathsf{A}$ and for all $F \in \mathcal{F}$. Wherever it is convenient, we assume that $S(x, \cdot)$ is locally bounded for all $x \in \mathsf{A}$. A functional $T\colon \mathcal{F} \to \mathsf{A}$ is *elicitable* if it possesses a strictly $\mathcal{F}$-consistent scoring function.

Definition 2.5 (Equivalence)**.** Two scoring function $S, \widetilde{S}\colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ are called *equivalent* if there is some $\mathcal{F}$-integrable function $a\colon \mathbb{R} \to \mathbb{R}$ and some $\lambda > 0$ such that $\widetilde{S}(x, y) = \lambda S(x, y) + a(y)$ for all $(x, y) \in \mathsf{A} \times \mathbb{R}$.

It is immediate that the above relation is indeed an equivalence relation. Moreover, if S and $\widetilde{S}$ are equivalent, then S is (strictly) $\mathcal{F}$-consistent for some functional T if and only if $\widetilde{S}$ is (strictly) $\mathcal{F}$-consistent for T.

Closely related to the concept of elicitability is the notion of *identifiability*.

Definition 2.6 (Identification functions and identifiability)**.** An $\mathcal{F}$-integrable map $V\colon \mathsf{A}\times \mathbb{R}\to\mathbb{R}^k$, where $\mathsf{A}\subseteq\mathbb{R}^k$, is an identification function for a functional $T\colon\mathcal{F}\to\mathsf{A}$ if $\bar{V}(T(F),F)=0$ for all $F\in\mathcal{F}$. It is a *strict* $\mathcal{F}$-identification function for T if it is an identification function and if $\bar{V}(x,F)=0$ implies that $x=T(F)$ for all $x\in\mathsf{A}$ and for all $F\in\mathcal{F}$. Wherever it is convenient, we assume that V is locally bounded *jointly* in both arguments. A functional $T\colon\mathcal{F}\to\mathsf{A}$ is *identifiable* if it possesses a strict $\mathcal{F}$-identification function.

Note that in contrast to Gneiting (2011) we assume that the functional T maps to A rather than to the power set of A.

For the sake of completeness, we list some assumptions used in Section 3 which were originally introduced in Fissler and Ziegel (2016) in the Appendix.

3 Elicitability and identifiability results

3.1 RVaR is not elicitable

It is well known that the mean-functional is elicitable with respect to the class $\mathcal{F}$ of probability distributions with finite mean. Value-at-Risk at level α is elicitable relative to the class $\mathcal{F}$ of probability distributions with unique α-quantiles.[1] Gneiting (2011) showed that expected shortfall (ES) fails to have convex level sets which implies that ES is not elicitable. On the other hand, Fissler and Ziegel (2016) provide a positive result showing that the pair $(\mathrm{VaR}_\alpha,\mathrm{ES}_\alpha)$ is elicitable. The following proposition treats the case of $\mathrm{RVaR}_{\alpha,\beta}$ for $0<\alpha<\beta<1$.

Proposition 3.1. *Let $0<\alpha<\beta<1$. If $\mathcal{F}$ contains all measures with finite support, the following assertions hold.*

(i) $(\mathrm{VaR}_\beta,\mathrm{RVaR}_{\alpha,\beta})\colon\mathcal{F}\to\mathbb{R}^2$ *does not have convex level sets.*

(ii) $(\mathrm{VaR}_\alpha,\mathrm{RVaR}_{\alpha,\beta})\colon\mathcal{F}\to\mathbb{R}^2$ *does not have convex level sets.*

(iii) $\mathrm{RVaR}_{\alpha,\beta}\colon\mathcal{F}\to\mathbb{R}$ *does not have convex level sets.*

Proof. We start with (i). Let $a<b<c<d$ where $a<2/3b+c<b$ and where $c\neq 0$. Define the two measures $F_1=\alpha\delta_a+\frac{1}{2}(\beta-\alpha)(\delta_b+\delta_c)+(1-\beta)\delta_d$ and $F_2=\frac{1}{4}(\alpha+3\beta)\delta_{(2/3b+d)}+\frac{1}{4}(\beta-\alpha)\delta_c+(1-\beta)\delta_d$. Then $\mathrm{VaR}_\beta(F_1)=\mathrm{VaR}_\beta(F_2)=\mathrm{VaR}_\beta(\frac{1}{2}(F_1+F_2))=c$, and $\mathrm{RVaR}_{\alpha,\beta}(F_1)=\mathrm{RVaR}_{\alpha,\beta}(F_2)=\frac{1}{2}(b+c)$. On the other hand, one obtains $\mathrm{RVaR}_{\alpha,\beta}(\frac{1}{2}(F_1+F_2))=\frac{1}{2}b+\frac{2}{3}c$, which shows (i). Assertion (ii) follows with a similar argument, whereas (iii) is a direct corollary of (i) or (ii). $\square$

[1] That is, if $\{x\in\mathbb{R}\mid \lim_{t\uparrow x}F(t)\le\alpha\le F(x)\}=\{\mathrm{VaR}_\alpha(F)\}$ for all $F\in\mathcal{F}$.

We would like to remark that Wang and Wei (2018, Example 7) provide an alternative proof that $\mathrm{RVaR}_{\alpha,\beta}$ does not have convex level sets for $0 < \alpha < \beta < 1$. Moreover, their Theorem 2 gives an alternative way of establishing assertions (i) and (ii) in Proposition 3.1.

Corollary 3.2. *Let $0 < \alpha < \beta < 1$. If $\mathcal{F}$ contains all measures with finite support, the following assertions hold.*

(i) $(\mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})\colon \mathcal{F} \to \mathbb{R}^2$ *is not elicitable.*

(ii) $(\mathrm{VaR}_\alpha, \mathrm{RVaR}_{\alpha,\beta})\colon \mathcal{F} \to \mathbb{R}^2$ *is not elicitable.*

(iii) $\mathrm{RVaR}_{\alpha,\beta}\colon \mathcal{F} \to \mathbb{R}$ *is not elicitable.*

Proof. This is a direct consequence of Proposition 3.1 and Theorem 6 in Gneiting (2011). □

With similar arguments one can show the assertions of Proposition 3.1 (and Corollary 3.2) if $\mathcal{F}$ contains all measures with compact support that are continuous with respect to the Lebesgue measure. With a continuity argument, one can extend this result to the class $\mathcal{F}$ containing mixtures of normal distributions.

3.2 RVaR is jointly elicitable with the corresponding quantiles

For any $\alpha \in (0,1)$, we define $S_\alpha\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}$, $S_\alpha(x,y) = (\mathbb{1}\{y \le x\} - \alpha)x - \mathbb{1}\{y \le x\}y$. Note that S_α is $\mathcal{F}$-consistent for VaR_α if $\int_{-\infty}^{x} |y| \,\mathrm{d}F(y) < \infty$ for all $F \in \mathcal{F}$ and all $x \in \mathbb{R}$. Moreover, it is strictly $\mathcal{F}$-consistent for VaR_α if all distributions in $\mathcal{F}$ have unique α-quantiles.

Now let $0 < \alpha < \beta < 1$ and consider the function $V\colon \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3$ defined as

$$V(x_1,x_2,x_3,y) = \begin{pmatrix} \mathbb{1}\{y \le x_1\} - \alpha \\ \mathbb{1}\{y \le x_2\} - \beta \\ x_3 + \frac{1}{\beta-\alpha}\big(S_\beta(x_2,y) - S_\alpha(x_1,y)\big) \end{pmatrix} \tag{3.1}$$

Using the notation $V = (V_1, V_2, V_3)^\top$, the important observation is that for any distribution F

$$\bar{V}_3(\mathrm{VaR}_\alpha(F), \mathrm{VaR}_\beta(F), x_3, F) = x_3 - \mathrm{RVaR}_{\alpha,\beta}(F). \tag{3.2}$$

This observation implies an identifiability result for the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ whose proof is simple and omitted.

Lemma 3.3. *Let $0 < \alpha < \beta < 1$. If $\mathcal{F}$ is a class of probability distributions such that $F(\mathrm{VaR}_\alpha(F)) = \alpha$ and $F(\mathrm{VaR}_\beta(F)) = \beta$ for all $F \in \mathcal{F}$, then the function V at (3.1) is an $\mathcal{F}$-identification function for the triplet $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$. If moreover the α- and β-quantiles are unique for all elements of $\mathcal{F}$, then V is a strict $\mathcal{F}$-identification function for T.*

Invoking the inequality at (2.3) the maximal sensible action domain for the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ is $\mathsf{A}_0 := \{(x_1,x_2,x_3) \in \mathbb{R}^3 \,|\, x_1 \le x_3 \le x_2\}$.

Theorem 3.4. *Let $\mathcal{F}$ be a class of distributions on $\mathbb{R}$, $0 < \alpha < \beta < 1$, and $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})\colon \mathcal{F} \to \mathsf{A} = \{(x_1,x_2,x_3) \in \mathbb{R}^3 \,|\, C_1 < x_1 \le x_3 \le x_2 < C_2\}$ where $-\infty \le C_1 < C_2 \le \infty$. Let $S\colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ be a scoring function of the form*

$$\begin{aligned} S(x_1,x_2,x_3,y) &= \big(\mathbb{1}\{y \le x_1\} - \alpha\big)g_1(x_1) - \mathbb{1}\{y \le x_1\}g_1(y) \\ &\quad + \big(\mathbb{1}\{y \le x_2\} - \beta\big)g_2(x_2) - \mathbb{1}\{y \le x_2\}g_2(y) \\ &\quad + \phi'(x_3)\Big(x_3 + \frac{1}{\beta-\alpha}\big(S_\beta(x_2,y) - S_\alpha(x_1,y)\big)\Big) - \phi(x_3) + a(y), \end{aligned} \tag{3.3}$$

where $a\colon \mathbb{R} \to \mathbb{R}$ is $\mathcal{F}$-integrable, $g_r\colon (C_1,C_2) \to \mathbb{R}$, $r \in \{1,2\}$, such that the functions $\mathbb{1}_{(-\infty,x_r]}g_r$ are $\mathcal{F}$-integrable for all $x_r \in (C_1,C_2)$, and $\phi\colon (C_1,C_2) \to \mathbb{R}$ is convex with subgradient $\phi'\colon (C_1,C_2) \to \mathbb{R}$. If for all $x_3 \in (C_1,C_2)$ the functions

$$G_{1,x_3}\colon (C_1,C_2) \to \mathbb{R}, \qquad x_1 \mapsto g_1(x_1) - x_1\phi'(x_3)/(\beta-\alpha), \tag{3.4}$$

$$G_{2,x_3}\colon (C_1,C_2) \to \mathbb{R}, \qquad x_2 \mapsto g_2(x_2) + x_2\phi'(x_3)/(\beta-\alpha) \tag{3.5}$$

are increasing, then S is $\mathcal{F}$-consistent for T. If moreover ϕ is strictly convex, the functions at (3.4) and (3.5) are strictly increasing, and any distribution in $\mathcal{F}$ has unique α- and β-quantiles, then S is strictly $\mathcal{F}$-consistent for T.

Proof. To simplify the notation in the proof, we shall occasionally evaluate the score on $(C_1,C_2)^3$ rather than on A. Let $(x_1,x_2,x_3) \in \mathsf{A}$, $F \in \mathcal{F}$ and $(t_1,t_2,t_3) := T(F)$. Then, since G_{1,x_3} is increasing, $(C_1,C_2) \times \mathbb{R} \ni (x_1',y) \mapsto S(x_1',x_2,x_3,y)$ is $\mathcal{F}$-consistent for VaR_α and it is strictly $\mathcal{F}$-consistent if G_{1,x_3} is strictly increasing and if the distributions in $\mathcal{F}$ have unique α-quantiles. Similar comments apply to the map $(C_1,C_2) \times \mathbb{R} \ni (x_2',y) \mapsto S(t_1,x_2',x_3,y)$. Hence,

$$\begin{aligned} 0 &\le \bar{S}(x_1,x_2,x_3,F) - \bar{S}(t_1,x_2,x_3,F) + \bar{S}(t_1,x_2,x_3,F) - \bar{S}(t_1,t_2,x_3,F) \\ &= \bar{S}(x_1,x_2,x_3,F) - \bar{S}(t_1,t_2,x_3,F) \end{aligned}$$

with a strict inequality under the conditions for strict consistency and if $(x_1,x_2) \ne (t_1,t_2)$. Finally,

$$\bar{S}(t_1,t_2,x_3,F) - \bar{S}(t_1,t_2,t_3,F) = \phi'(x_3)(x_3 - t_3) - \phi(x_3) + \phi(t_3) \ge 0, \tag{3.6}$$

since ϕ is convex. If ϕ is strictly convex and if $x_3 \ne t_3$, then the inequality in (3.6) is strict. ☐

Remark 3.5. (i) If $C_1, C_2 \in \mathbb{R}$, then Theorem 3.4 holds also for the action domain $\mathsf{A} = \{(x_1,x_2,x_3) \in \mathbb{R}^3 \,|\, C_1 \le x_1 \le x_3 \le x_2 \le C_2\}$.

(ii) Even though the maximal sensible action domain for $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ is $\mathsf{A}_0 := \{(x_1,x_2,x_3) \in \mathbb{R}^3 \,|\, x_1 \le x_3 \le x_2\}$, the proof of Theorem 3.4 shows that the scoring function given at (3.3) is even strictly $\mathcal{F}$-consistent on the Cartesian product $\mathsf{A}_1' \times \mathsf{A}_2' \times \mathsf{A}_3'$ where $\mathsf{A}_r' := \{x_r \in \mathbb{R} \,|\, \exists (z_1,z_2,z_3) \in \mathsf{A},\ z_r = x_r\}$, $r \in \{1,2,3\}$, is the projection of A to the rth component. This enables the evaluation of forecasts ignoring the crucial inequality at (2.3).

(iii) If the scoring S is of the form at (3.3) such that ϕ is strictly convex and the functions G_1 and G_2 are strictly increasing, but some distributions $F \in \mathcal{F}$ fail to have unique α- or β-quantiles, then S also fails to be strictly $\mathcal{F}$-consistent. However, it is still strictly $\mathcal{F}$-consistent in the RVaR-component. That is, for $F \in \mathcal{F}$

$$\arg\min_{x \in \mathsf{A}_0} \bar{S}(x, F) = q_\alpha(F) \times q_\beta(F) \times \{\mathrm{RVaR}_{\alpha,\beta}(F)\},$$

where $q_\gamma(F) := \{y \in \mathbb{R} \mid \lim_{t \uparrow y} F(t) \leq \gamma \leq F(y)\}$ is the full set-valued γ-quantile.

Making use of the relation at (2.2) and the revelation principle (Osband, 1985; Gneiting, 2011; Fissler, 2017), Theorem 3.4 establishes that the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_{1-\alpha}, W_\alpha)$ is elicitable where W_α is the α-Winsorized mean. Moreover, it gives a rich class of strictly consistent scoring function for this triplet. The following proposition is useful to construct examples; see Section 6.

Proposition 3.6. *Let $0 < \alpha < \beta < 1$ and $-\infty \leq C_1 < C_2 \leq \infty$. Let S be a scoring function of the form (3.3) with a (strictly) convex function ϕ, and functions g_1, g_2 satisfying the conditions (3.4) and (3.5).*

(i) *The subgradient ϕ' of ϕ is necessarily bounded.*

(ii) *S is equivalent to a scoring function $\tilde{S}$ of the form (3.3) with a (strictly) convex function $\tilde{\phi}$ such that $\tilde{\phi}'$ is bounded with $-\inf_{x \in (C_1,C_2)} \tilde{\phi}'(x) = \sup_{x \in (C_1,C_2)} \tilde{\phi}'(x) = \beta - \alpha$, and strictly increasing functions $\tilde{g}_1$, $\tilde{g}_2$ such that their one-sided derivatives are bounded below by one.*

Proof. (i) The proof is similar to the one of Corollary 5.5 in Fissler and Ziegel (2016). Take some $x_1, x_1' \in \mathsf{A}_1'$ with $x_1 < x_1'$. Then, for any $x_3 \in \mathsf{A}_3'$ one obtains $\phi'(x_3)/(\beta - \alpha) \leq (g_1(x_1) - g_1(x_1'))/(x_1 - x_1') < \infty$. One obtains that $\sup_{x_3 \in \mathsf{A}_3'} \phi'(x_3) < \infty$. With similar arguments one can show that $\inf_{x_3 \in \mathsf{A}_3'} \phi'(x_3) > -\infty$.

(ii) For any $c \in \mathbb{R}$, if we replace ϕ with $\widehat{\phi} : x \mapsto \phi(x) + cx$, g_1 with $\widehat{g}_1 : x \mapsto g_1(x) + cx/(\beta - \alpha)$, and g_2 with $\widehat{g}_2 : x \mapsto g_2(x) + cx/(\beta - \alpha)$ in the formula (3.3) for S, then S does not change and $\widehat{\phi}$ is (strictly) convex if and only if ϕ is (strictly) convex. Furthermore, conditions (3.4) and (3.5) hold for ϕ, g_1, g_2 if and only if they hold for $\widehat{\phi}$, $\widehat{g}_1$ and $\widehat{g}_2$. By part (i) of the proposition ϕ' is bounded. Therefore, we can assume without loss of generality that $-\inf_{x \in (C_1,C_2)} \phi'(x) = \sup_{x \in (C_1,C_2)} \phi'(x) = c_0$. By scaling of the scoring function S, we obtain an equivalent scoring function where we can assume that $c_0 = \beta - \alpha$.

Let $x, x' \in (C_1, C_2)$ such that $x < x'$. Condition (3.4) implies that

$$\frac{g_1(x') - g_1(x)}{x' - x} \geq \frac{1}{\beta - \alpha} \sup_{x_3 \in (C_1,C_2)} \phi'(x_3) = 1.$$

In particular, g_1 is strictly increasing, and therefore, one-sided derivatives exist everywhere and by the above inequality they are bounded below by one. The argument for g_2 works analogously.

∎

Using Osband's principle (Fissler and Ziegel, 2016, Theorem 3.2), one can also establish a necessary condition for strict consistency of scoring functions for the triplet $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$. For any $x_3 \in \mathsf{A}'_3$ and $m \in \{1,2\}$, let $\mathsf{A}'_{m,x_3} := \{x_m \in \mathbb{R} \,|\, \exists (z_1, z_2, z_3) \in \mathsf{A},\ z_m = x_m,\ z_3 = x_3\}$.

Theorem 3.7. *Let $\mathcal{F}$ be a class of continuously differentiable distributions on $\mathbb{R}$ and $0 < \alpha < \beta < 1$. Assume that all distributions in $\mathcal{F}$ have unique α- and β-quantiles. Let $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})\colon \mathcal{F} \to \mathsf{A} \subseteq \mathsf{A}_0$. Then, V defined at (3.1) is a strict $\mathcal{F}$-identification function for T which satisfies Assumption (V3). If Assumptions (V1), and (F1) hold and $(V_1, V_2)^\top$ satisfies Assumption (V4), then any strictly $\mathcal{F}$-consistent scoring function $S\colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ for T that satisfies assumptions (VS1) and (S2) is necessarily of the form given at (3.3) almost everywhere, where the functions $G_{r,x_3}\colon \mathsf{A}'_{r,x_3} \to \mathbb{R}$, $r \in \{1,2\}$, $x_3 \in \mathsf{A}'_3$, at (3.4) and (3.5) are strictly increasing and $\phi\colon \mathsf{A}'_3 \to \mathbb{R}$ is strictly convex.*

Proof. The first assertion of the theorem follows from the identity at (3.2) and Assumption (V3) is satisfied since all $F \in \mathcal{F}$ are assumed to be continuously differentiable.

Let $F \in \mathcal{F}$ with derivative f and let $x \in \mathrm{int}(\mathsf{A})$. Then one obtains

$$\bar{V}_3(x, F) = x_3 + \frac{1}{\beta - \alpha}\left(x_2(F(x_2) - \beta) - x_1(F(x_1) - \alpha) - \int_{x_1}^{x_2} y f(y)\,\mathrm{d}y\right)$$

The partial derivatives of V are given by

$$\partial_m \bar{V}_1(x, F) = \begin{cases} f(x_1), & m = 1 \\ 0, & m = 2 \\ 0, & m = 3, \end{cases} \qquad \partial_m \bar{V}_2(x, F) = \begin{cases} 0, & m = 1 \\ f(x_2), & m = 2 \\ 0, & m = 3, \end{cases}$$

$$\partial_m \bar{V}_3(x, F) = \begin{cases} -(F(x_1) - \alpha)/(\beta - \alpha), & m = 1 \\ (F(x_2) - \beta)/(\beta - \alpha), & m = 2 \\ 1, & m = 3. \end{cases}$$

An adaptation of Osband's Principle (Fissler and Ziegel, 2016, Theorem 3.2) yields the existence of continuously differentiable functions $h_{lm}\colon \mathrm{int}(\mathsf{A}) \to \mathbb{R}$, $l, m \in \{1,2,3\}$, such that for $m \in \{1,2,3\}$

$$\partial_m \bar{S}(x, F) = \sum_{i=1}^{3} h_{mi}(x) \bar{V}_i(x, F).$$

Since we assume that $\bar{S}(\cdot, F)$ is twice continuously differentiable for any $F \in \mathcal{F}$, the second order partial derivatives need to commute. Let $t = T(F)$. Then $\partial_1 \partial_2 \bar{S}(t, F) = \partial_2 \partial_1 \bar{S}(t, F)$ is equivalent to

$$h_{21}(t) f(t_1) = h_{12}(t) f(t_2).$$

This needs to hold for all $F \in \mathcal{F}$. The variation in the densities implied by Assumption (V4) in combination with the surjectivity of T yield that $h_{12} = h_{21} = 0$ on $\mathrm{int}(\mathsf{A})$. Similarly, evaluating $\partial_1\partial_3\bar{S}(x,F) = \partial_3\partial_1\bar{S}(x,F)$ and $\partial_2\partial_3\bar{S}(x,F) = \partial_3\partial_2\bar{S}(x,F)$ at $x = t = T(F)$ yields

$$h_{13}(t) = h_{31}(t)f(t_1), \qquad h_{23}(t) = h_{32}(t)f(t_2)\,.$$

Using again Assumption (V4) as well as the surjectivity of T, this implies that

$$h_{13} = h_{31} = h_{23} = h_{32} = 0\,.$$

So we are left with characterising h_{mm} for $m \in \{1,2,3\}$. Note that Assumption (V1) implies that for any $x = (x_1,x_2,x_3) \in \mathrm{int}(\mathsf{A})$ there are two distributions $F_1, F_2 \in \mathcal{F}$ such that $(F_1(x_1)-\alpha, F_1(x_2)-\beta)^\top$ and $(F_2(x_1)-\alpha, F_2(x_2)-\beta)^\top$ are linearly independent. Then, the requirement that

$$\partial_1\partial_2\bar{S}(x,F) = \partial_1 h_{22}(x)(F(x_2)-\beta) = \partial_2 h_{11}(x)(F(x_1)-\alpha) = \partial_2\partial_1\bar{S}(x,F)$$

for all $x \in \mathrm{int}(\mathsf{A})$ and for all $F \in \mathcal{F}$ implies that $\partial_1 h_{22} = \partial_2 h_{11} = 0$.
Starting with $\partial_1\partial_3\bar{S}(x,F) = \partial_3\partial_1\bar{S}(x,F)$, implies that

$$\partial_1 h_{33}\bar{V}_3(x,F) = \big(\partial_3 h_{11}(x) + h_{33}(x)/(\beta-\alpha)\big)\bar{V}_1(x,F).$$

Again, Assumption (V1) implies that there are $F_1, F_2 \in \mathcal{F}$ such that $\big(\bar{V}_1(x,F_1), \bar{V}_3(x,F_1)\big)^\top$ and $\big(\bar{V}_1(x,F_2), \bar{V}_3(x,F_2)\big)^\top$ are linearly independent. Hence, we obtain that $\partial_1 h_{33} = 0$ and $\partial_3 h_{11} = -h_{33}/(\beta-\alpha)$. With the same argumentation and starting from $\partial_2\partial_3\bar{S}(x,F) = \partial_3\partial_2\bar{S}(x,F)$ one can show that $\partial_2 h_{33} = 0$ and $\partial_3 h_{22} = h_{33}/(\beta-\alpha)$. That means there exist functions $c_1\colon \{(x_1,x_3) \in \mathbb{R}^2 \,|\, \exists (z_1,z_2,z_3) \in \mathrm{int}(\mathsf{A}),\ x_1 = z_1, x_3 = z_3\} \to \mathbb{R}$, $c_2\colon \{(x_2,x_3) \in \mathbb{R}^2 \,|\, \exists (z_1,z_2,z_3) \in \mathrm{int}(\mathsf{A}),\ x_2 = z_2, x_3 = z_3\} \to \mathbb{R}$ and $c_3\colon \mathrm{int}(\mathsf{A})'_3 \to \mathbb{R}$ and some $z \in \mathrm{int}(\mathsf{A})'_3$ such that for any $x = (x_1,x_2,x_3) \in \mathrm{int}(\mathsf{A})$

$$\begin{aligned}
h_{33}(x) &= c_3(x_3),\\
h_{11}(x) &= c_1(x_1,x_3) = -\frac{1}{\beta-\alpha}\int_z^{x_3} c_3(z)\,\mathrm{d}z + b_1(x_1),\\
h_{22}(x) &= c_2(x_2,x_3) = \frac{1}{\beta-\alpha}\int_z^{x_3} c_3(z)\,\mathrm{d}z + b_2(x_2),
\end{aligned}$$

where $b_r\colon\ \mathrm{int}(\mathsf{A})'_r \to \mathbb{R}$, $r \in \{1,2\}$. Due to the fact that any component of T is mixture-continuous[2] and since $\mathcal{F}$ is convex and T surjective, the projection $\mathrm{int}(\mathsf{A})'_3$ is an open interval. Hence, $[\min(z,x_3), \max(z,x_3)] \subset \mathrm{int}(\mathsf{A})'_3$. Due to Assumptions (V3) and (S2), Theorem 3.2 in Fissler and Ziegel (2016) implies that c_1, c_2, c_3 are locally Lipschitz continuous. The above calculations imply that the Hessian of the expected score, $\nabla^2\bar{S}(x,F)$, at its minimiser $x = t = T(F)$ takes the form

$$\nabla^2\bar{S}(t,F) = \begin{pmatrix} c_1(t_1,t_3)f(t_1) & 0 & 0\\ 0 & c_2(t_2,t_3)f(t_2) & 0\\ 0 & 0 & c_3(t_3)\end{pmatrix}.$$

[2]For convex $\mathcal{F}$ a functional $T\colon \mathcal{F} \to \mathbb{R}^k$ is called mixture-continuous if for any $F, G \in \mathcal{F}$ the map $[0,1] \ni \lambda \mapsto T((1-\lambda)F + \lambda G)$ is continuous.

Since t is a minimiser of the expected score, the Hessian must be positive semi-definite. Invoking the surjectivity of T once again, this shows that $c_1, c_2, c_3 \geq 0$. More to the point, invoking the continuous differentiability of the expected score and the fact that S is strictly $\mathcal{F}$-consistent for T one obtains that for any $F \in \mathcal{F}$ with $t = T(F)$ and for any $v \in \mathbb{R}^3$, $v \neq 0$, there exists an $\varepsilon > 0$ such that

$$\frac{\mathrm{d}}{\mathrm{d}s}\bar{S}(t + sv, F) \begin{cases} < 0, & \forall s \in (-\varepsilon, 0) \\ = 0, & s = 0 \\ > 0, & \forall s \in (0, \varepsilon). \end{cases} \tag{3.7}$$

For $v = e_3 = (0,0,1)^\top$, (3.7) means that for any $F \in \mathcal{F}$ with $t = T(F)$ there is an $\varepsilon > 0$ such that

$$\frac{\mathrm{d}}{\mathrm{d}s}\bar{S}(t + se_3, F) = c_3(t_3 + s)s \begin{cases} < 0, & \forall s \in (-\varepsilon, 0) \\ = 0, & s = 0 \\ > 0, & \forall s \in (0, \varepsilon). \end{cases}$$

That means $c_3(t_3 + s) > 0$ for all $s \in (-\varepsilon, \varepsilon) \setminus \{0\}$. Using the surjectivity of T and invoking a compactness argument, c_3 attains a 0 only finitely many times on any compact interval. Recall that $\mathrm{int}(\mathsf{A})'_3$ is an open interval. Hence, it can be approximated by an increasing sequence of compact intervals. Therefore, $c_3^{-1}(\{0\})$ is at most countable and therefore a Lebesgue null set. With similar arguments one can show that for any $x_3 \in \mathrm{int}(\mathsf{A})'_3$, the sets $\{x_1 \in \mathbb{R} \,|\, \exists (z_1, z_2, z_3) \in \mathrm{int}(\mathsf{A}),\ x_1 = z_1,\ x_3 = z_3,\ c_1(x_1, x_3) = 0\}$ and $\{x_2 \in [x_3, \infty) \,|\, \exists (z_1, z_2, z_3) \in \mathrm{int}(\mathsf{A}),\ x_2 = z_2,\ x_3 = z_3,\ c_2(x_2, x_3) = 0\}$ are at most countable and therefore also Lebesgue null sets.

Finally, using Proposition 1.1 in Fissler and Ziegel (2019a) one obtains that S is almost everywhere of the form (3.3). Moreover, it holds almost everywhere that $\phi'' = c_3$ and $g'_m = b_m$ for $m \in \{1, 2\}$. Hence, ϕ is strictly convex and the functions at (3.4) and (3.5) are strictly increasing. □

Combining Theorems 3.4 and 3.7, one can show that the scoring functions given at (3.3) are essentially the only strictly consistent scoring functions for the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ on the action domain $\mathsf{A} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, C_1 < x_1 \leq x_3 \leq x_2 < C_2\}$.

Corollary 3.8. *Let the conditions of Theorem 3.7 prevail and let* $\mathsf{A} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, C_1 < x_1 \leq x_3 \leq x_2 < C_2\}$ *for some* $-\infty \leq C_1 < C_2 \leq \infty$*. Then, a scoring function* $S \colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ *is strictly* $\mathcal{F}$*-consistent for* $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$, $0 < \alpha < \beta < 1$, *if and only if it is of the form given at* (3.3) *and with the conditions specified around* (3.4) *and* (3.5)*. Moreover, the function* $\phi' \colon (C_1, C_2) \to \mathbb{R}$ *is necessarily bounded.*

Proof. For the proof it suffices to show that G_{r,x_3} is not only increasing on A'_{r,x_3} for any $x_3 \in \mathsf{A}'_3$ but on $\mathsf{A}'_r = (C_1, C_2)$, $r \in \{1, 2\}$. For $x_3 \in (C_1, C_3) = \mathsf{A}'_3$, we have $\mathsf{A}'_{1,x_3} = (C_1, x_3]$ and $\mathsf{A}'_{2,x_3} = [x_3, C_2)$. Let $x_3 \in \mathsf{A}'_3$ and $x_1, x'_1 \in \mathsf{A}'_1$ with $x_1 < x'_1$. If

$x_1, x_1' \in \mathsf{A}'_{1,x_3}$ there is nothing to show. If however $x_3 < x_1'$, then $x_1, x_1' \in \mathsf{A}'_{1,x_1'}$. This means that

$$\begin{aligned} 0 &\le g_1(x_1') - g_1(x_1) - (x_1' - x_1)\phi'(x_1')/(\beta - \alpha) \\ &\le g_1(x_1') - g_1(x_1) - (x_1' - x_1)\phi'(x_3)/(\beta - \alpha) \end{aligned}$$

where the second inequality stems from the fact that ϕ' is increasing. If the function $G_{1,x_1'}$ is strictly increasing, then the first inequality is strict. The argument for G_{2,x_3} works analogously. □

Similarly to Remark 3.5(i), Corollary 3.8 works also on the action domain $\mathsf{A} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, C_1 \le x_1 \le x_3 \le x_2 \le C_2\}$ if $C_1, C_2 \in \mathbb{R}$. Besides the maximal action domain $\mathsf{A}_0 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 \le x_3 \le x_2\}$ the practically most relevant choices of action domains in Corollary 3.8 are

$$\begin{aligned} \mathsf{A}_0^+ &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, 0 < x_1 \le x_3 \le x_2\}, \\ \mathsf{A}_0^- &= \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 \le x_3 \le x_2 < 0\}. \end{aligned}$$

Concrete examples for choices of the functions g_1, g_2, and ϕ for the scoring function S at (3.3) are given and discussed in Section 6.

Remark 3.9. Note the structural difference of Theorems 3.4 and 3.7 to Frongillo and Kash (2015, Theorem 1 and Corollary 1), Brehmer (2017, Proposition 4.14) and in particular Fissler and Ziegel (2016, Theorem 5.2 and Corollary 5.5). Our functional of interest, $\mathrm{RVaR}_{\alpha,\beta}$ with $0 < \alpha < \beta < 1$, is not a minimum of a scoring function, but a difference of minima of two scoring functions. Indeed, while $\mathrm{ES}_\beta(F) = -\frac{1}{\beta}\bar{S}_\beta(\mathrm{VaR}_\beta(F), F)$, we have that

$$\mathrm{RVaR}_{\alpha,\beta}(F) = -\frac{1}{\beta - \alpha}\big(\bar{S}_\beta(\mathrm{VaR}_\beta(F), F) - \bar{S}_\alpha(\mathrm{VaR}_\alpha(F), F)\big)\,.$$

This structural difference is reflected in the minus sign appearing at (3.4). In particular, it means that the functions g_1 and g_2 cannot identically vanish if we want to ensure strict consistency of S whereas the corresponding functions in Theorem 5.2 in Fissler and Ziegel (2016) may well be set to zero.

4 Translation invariance and homogeneity

There are many choices for the functions g_1, g_2, and ϕ appearing in the formula for the scoring function S at (3.3). Often, these choices can be limited by impoing secondary desirable criteria on S. In this section we show that, unfortunately, standard criteria such as translation invariance and homogeneity are not fruitful for RVaR.

If one is interested in scoring functions with an action domain of the form $\mathsf{A} = \{x \in \mathbb{R}^3 \,|\, C_1 < x_1 \le x_3 \le x_2 < C_2\}$ possessing the additional property of translation invariant score differences, the only sensible choice is $C_1 = -\infty$, $C_2 = \infty$, amounting to the maximal action domain A_0. Similarly, for scoring functions with positively homogeneous score differences, the most interesting choices for action domains are $\mathsf{A} \in \{\mathsf{A}_0, \mathsf{A}_0^+, \mathsf{A}_0^-\}$.

Proposition 4.1 (Translation invariance)**.** *Let $0 < \alpha < \beta < 1$. Under the conditions of Theorem 3.7 there are no strictly $\mathcal{F}$-consistent scoring functions for $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ on A_0 with translation invariant score differences.*

Proof. Using Theorem 3.7 any strictly $\mathcal{F}$-consistent scoring function for T must be of the form at (3.3) where in particular ϕ is strictly convex, twice differentiable, and ϕ' is bounded, invoking Proposition 3.6. Assume that S has translation invariant score differences. That means that the function $\Psi\colon \mathbb{R} \times \mathsf{A}_0 \times \mathsf{A}_0 \times \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned}\Psi(z, x, x', y) = S(x_1 + z, x_2 + z, x_3 + z, y + z) - S(x_1' + z, x_2' + z, x_3' + z, y + z) \\ - S(x_1, x_2, x_3, y) + S(x_1', x_2', x_3', y)\end{aligned}$$

vanishes. Then, for all $x \in \mathsf{A}_0$ and for all $z, y \in \mathbb{R}$

$$0 = \frac{\mathrm{d}}{\mathrm{d}x_3}\Psi(z, x, x', y) = \big(\phi''(x_3 + z) - \phi''(x_3)\big)\Big(x_3 + \frac{1}{\beta - \alpha}\big(S_\beta(x_2, y) - S_\alpha(x_1, y)\big)\Big).$$

Therefore, ϕ'' is constant. Since ϕ is convex and that means that $\phi'(x_3) = dx_3 + d'$ with $d > 0$. But since $\mathsf{A}_3' = \mathbb{R}$, ϕ' is unbounded, which is a contradiction. □

The proof of Proposition 4.1 closely follows the one in Proposition 4.10 in Fissler and Ziegel (2019b). The fact that the latter assertion entails a positive result has the following background: The strictly consistent scoring function for $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ in Proposition 4.10 in Fissler and Ziegel (2019b) works only on a very restricted action domain. To guarantee strict consistency on such an action domain, one would need a refinement of Theorem 3.4 in the spirit of Proposition 2.1 in Fissler and Ziegel (2019a). However, since such a positive result on a quite restricted action domain is practically irrelevant, we dispense with such a refinement and only state the relevant negative result here.

Proposition 4.2 (Homogeneity)**.** *Let $0 < \alpha < \beta < 1$. Under the conditions of Theorem 3.7 there are no strictly $\mathcal{F}$-consistent scoring functions for $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ on $\mathsf{A} \in \{\mathsf{A}_0, \mathsf{A}_0^+, \mathsf{A}_0^-\}$ with positively homogeneous score differences.*

Proof. Using Theorem 3.7 any strictly $\mathcal{F}$-consistent scoring function for T must be of the form at (3.3) where in particular ϕ is strictly convex, twice differentiable, and ϕ' is bounded, invoking Proposition 3.6. Assume that S has positively homogeneous score differences of some degree $b \in \mathbb{R}$. That means that the function $\Psi\colon (0,\infty) \times \mathsf{A} \times \mathsf{A} \times \mathbb{R} \to \mathbb{R}$,

$$\Psi(c, x, x', y) = S(cx, cy) - S(cx', cy) - c^b S(x, y) + c^b S(x', y)$$

vanishes. Therefore, for all $x \in \mathsf{A}$ and for all $y \in \mathbb{R}$ and all $c > 0$

$$0 = \frac{\mathrm{d}}{\mathrm{d}x_3}\Psi(z, x, x', y) = \big(c^2\phi''(cx_3) - c^b\phi''(x_3)\big)\Big(x_3 + \frac{1}{\beta - \alpha}\big(S_\beta(x_2, y) - S_\alpha(x_1, y)\big)\Big). \tag{4.1}$$

For the sake of brevity, we only consider the case $\mathsf{A} = \mathsf{A}_0^-$, the other cases being similar. Equation (4.1) implies that that $\phi''(-x_3) = \phi''(-1)x_3^{b-2}$ for any $x_3 > 0$. Due to the strict convexity of ϕ, we need that $\phi''(-1) > 0$. However, for $b \geq 1$, $\inf_{x_3>0} \phi'(-x_3) = -\infty$ and for $b \leq 1$, $\sup_{x_3>0} \phi'(-x_3) = \infty$. Hence, ϕ' cannot be bounded. □

Remark 4.3. The negative result of Proposition 4.2 should be compared with the results of Theorem C.3 in Nolde and Ziegel (2017) characterising homogeneous strictly consistent scoring functions for the pair $(\mathrm{VaR}_\beta, \mathrm{ES}_\beta)$. Since they use a different sign convention for VaR and ES than we do in this paper, their choice of the action domain $\mathbb{R} \times (0, \infty)$ corresponds to our choice A_0^-. Note that when interpreting $\mathrm{RVaR}_{\alpha,\beta}$ as a risk measure, negative values of RVaR are the more interesting and relevant ones. Inspecting the proof of Proposition 4.2 and part (i) of Proposition 3.6 one makes the following observation: For $b \geq 1$, Nolde and Ziegel (2017) state an impossibility result for their choice of action domain. In fact, the problem occurring in our context is that ϕ' is not bounded *from below*. In Proposition 3.6 this property is implied by the fact that the function G_{2,x_3} at (3.5) is increasing. And it is exactly such a condition that is also present for strictly consistent scoring functions for the pair $(\mathrm{VaR}_\beta, \mathrm{ES}_\beta)$; see Theorem 5.2 in Fissler and Ziegel (2016). On the other hand, the complication for $b < 1$ stems from the fact that ϕ' is not bounded *from above*. This condition is related to the monotonicity of G_{1,x_3} at (3.4). Such a condition is not present for strictly consistent scoring functions for the pair $(\mathrm{VaR}_\beta, \mathrm{ES}_\beta)$. Correspondingly, there can be homogeneous and strictly consistent scoring functions for $b < 1$ for this pair (Nolde and Ziegel, 2017) while this is not possible for the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$.

5 Mixture representation of scoring functions

When forecasts are compared and ranked with respect to a consistent scoring functions, one has to be aware that in the presence of non-nested information sets, model misspecification and/or finite samples, the ranking may depend on the chosen consistent scoring functions (Patton, 2019). In the specific case of RVaR, this means that the forecast ranking may depend on the specific choice for the functions g_1, g_2, and ϕ appearing in Theorem 3.4. A possible remedy to this problem is to compare forecasts simultaneously with respect to all consistent scoring functions in terms of Murphy diagrams as introduced by Ehm et al. (2016). Murphy diagrams are based on the fact that the class of all consistent scoring functions can be characterized as a class of mixtures of elementary scoring functions that depend on a low-dimensional parameter. The following theorem provides such a mixture representation for the scoring functions at (3.3). Recall that $S_\alpha(x, y) = (\mathbb{1}\{y \leq x\} - \alpha)x - \mathbb{1}\{y \leq x\}y$.

Theorem 5.1. *Let* $0 < \alpha < \beta < 1$, $\mathsf{A}_0 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 \leq x_3 \leq x_2\}$. *Any scoring function* $S : \mathsf{A}_0 \times \mathbb{R} \to \mathbb{R}$ *of the form at* (3.3) *with* $a\colon \mathbb{R} \to \mathbb{R}$ *chosen such that*

$S(y,y,y,y)=0$ can be written as

$$S(x_1,x_2,x_3,y) = \int L_v^1(x_1,y)\,\mathrm{d}H_1(v) + \int L_v^2(x_2,y)\,\mathrm{d}H_2(v) + \int L_v^3(x_1,x_2,x_3,y)\,\mathrm{d}H_3(v), \tag{5.1}$$

where

$$\begin{aligned} L_v^1(x_1,y) &= (\mathbb{1}\{y \le x_1\} - \alpha)(\mathbb{1}\{v \le x_1\} - \mathbb{1}\{v \le y\}) \\ L_v^2(x_2,y) &= (\mathbb{1}\{y \le x_2\} - \beta)(\mathbb{1}\{v \le x_2\} - \mathbb{1}\{v \le y\}) \\ L_v^3(x_1,x_2,x_3,y) &= \frac{1}{\beta-\alpha}\Big(\mathbb{1}\{v > x_3\}(S_\alpha(x_1,y)+\alpha y) + \mathbb{1}\{v \le x_3\}(S_\beta(x_2,y)+\beta y)\Big) \\ &\quad + (\mathbb{1}\{v \le x_3\} - \mathbb{1}\{v \le y\})v, \end{aligned}$$

and H_1, H_2 are locally finite measures on $\mathbb{R}$, H_3 is a finite measure on $\mathbb{R}$. If H_3 puts positive mass on all open intervals, then S is strictly consistent. Conversely, for any choice of measures H_1, H_2, H_3 with the above restrictions, we obtain a scoring function of the form (3.3).

Proof. An increasing function $h : \mathbb{R} \to \mathbb{R}$ can always be written as

$$h(x) = \int (\mathbb{1}\{v \le x\} - \mathbb{1}\{v \le z\})\,\mathrm{d}H(v) + C, \quad x \in \mathbb{R}, \tag{5.2}$$

for some locally finite measure H, and $z \in \mathbb{R} \cup \{\pm\infty\}$, $C \in \mathbb{R}$. The function h is strictly increasing if and only if H puts positive mass on all open intervals. Furthermore, the one-sided derivatives of h are bounded below by $\lambda > 0$ if and only if $H(A) \ge \lambda\mathcal{L}(A)$ for all Borel sets $A \subseteq \mathbb{R}$.

Let ϕ satisfy the conditions of Proposition 3.6 (ii). Then, we can find a measure H_3 as stated in the theorem such that for all $x_3 \in (C_1, C_2)$, we have

$$\phi'(x_3) = \int \mathbb{1}\{v \le x_3\}\,\mathrm{d}H_3(v) - \lambda(\beta-\alpha) = \int \left(\mathbb{1}\{v \le x_3\} - \frac{1}{2}\right)\mathrm{d}H_3(v).$$

Using Fubini's theorem, we find that for some $z \in (C_1, C_2)$

$$\begin{aligned} \phi(x_3) - \phi(y) &= \int (\mathbb{1}\{w \le x_3\} - \mathbb{1}\{w \le z\})\phi'(w) - (\mathbb{1}\{w \le y\}) - \mathbb{1}\{w \le z\})\phi'(w)\,\mathrm{d}w \\ &= \int (\mathbb{1}\{w \le x_3\} - \mathbb{1}\{w \le y\}) \int \left(\mathbb{1}\{v \le w\} - \frac{1}{2}\right)\mathrm{d}H_3(v)\,\mathrm{d}w \\ &= \int\int (\mathbb{1}\{w \le x_3\} - \mathbb{1}\{w \le y\})\mathbb{1}\{v \le w\}\mathrm{d}w\,\mathrm{d}H_3(v) - \int \frac{1}{2}(x_3 - y)\,\mathrm{d}H_3(v) \\ &= \int \mathbb{1}\{v \le x_3\}(x_3 - v) - \mathbb{1}\{v \le y\}(y - v) - \frac{1}{2}(x_3 - y)\,\mathrm{d}H_3(v). \end{aligned}$$

Using (3.3), (5.2) and Proposition 3.6 it is straight forward to check that a scoring function of the form (3.3) can be written as in (5.1) with L_v^3 replaced by

$$\tilde{L}_v^3(x_1, x_2, x_3, y) = \left(\mathbb{1}\{v \le x_3\} - \frac{1}{2}\right)\left(x_3 + \frac{1}{\beta - \alpha}(S_\beta(x_2, y) - S_\alpha(x_1, y))\right) - \frac{1}{2}|x_3 - v| + \frac{1}{2}|y - v|,$$

locally finite measures $\tilde{H}_1$, $\tilde{H}_2$ instead of H_1, H_2 such that $\tilde{H}_i(A) \ge \lambda\mathcal{L}(A)$ for $i = 1, 2$, some $\lambda > 0$ and all Borel sets $A \subseteq \mathbb{R}$, and a finite measure H_3 with $H_3(\mathbb{R}) = 2\lambda(\beta - \alpha)$. We can write $\tilde{H}_i = H_i + \lambda\mathcal{L}$, $i = 1, 2$, for some locally finite measures H_i, $i = 1, 2$. Integrating $v \mapsto L_v^1$ with respect to $\lambda\mathcal{L}$, we obtain the function $\lambda(S_\alpha(x_1, y) + \alpha y)$, and analogously for L_v^2. Using that $H_3(\mathbb{R})$ is a finite measure with total mass $2\lambda(\beta - \alpha)$ yields the claim with

$$\begin{aligned} L_v^3(x_1, x_2, x_3, y) = &\frac{1}{2(\beta - \alpha)}(S_\beta(x_2, y) + \beta y + S_\alpha(x_1, y) + \alpha y) \\ &+ \left(\mathbb{1}\{v \le x_3\} - \frac{1}{2}\right)\left(x_3 + \frac{1}{\beta - \alpha}(S_\beta(x_2, y) - S_\alpha(x_1, y))\right) \\ &- \frac{1}{2}|x_3 - v| + \frac{1}{2}|y - v|, \end{aligned}$$

which is equal to the formula given in the statement of the theorem. The scoring functions L_v^1 and L_v^2 are consistent for VaR at level α and β, respectively. The scoring function L_v^3 is of the form at (3.3) with $g_1(x) = g_2(x) = x/(2\beta - 2\alpha)$ and $\phi(x) = |x - v|/2$, which renders it a consistent scoring function for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$. □

6 Simulations

Due to the negative results in Section 4 it is challenging to suggest concrete examples for the choices of the functions ϕ, g_1 and g_2 in (3.3). In Table 1, we give some first suggestions. The scoring function S_4 is in the spirit of the Huber loss. It is only strictly consistent on $\mathsf{A} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, c_1 < x_1 \le x_3 \le x_2 < c_2\}$ but remains consistent for all of A_0. We illustrate the discrimination ability of the suggested scoring functions with a slightly extended version of a simulation example of Gneiting et al. (2007) which has also been considered in Fissler et al. (2016).

Let $(\mu_t)_{t=1,\dots,N}$ be a sequence of independent standard normal random variables. Conditional on μ_t, Y_t is normally distributed with mean μ_t and variance 1, which we denote by $\mathcal{N}(\mu_t, 1)$. The first forecaster has access to μ_t and uses the correct conditional distribution for prediction, that is, she predicts

$$f_t = (f_{1,t}, f_{2,t}, f_{3,t}) = \left(\mu_t + \Phi^{-1}(\alpha), \mu_t + \Phi^{-1}(\beta), \mu_t - \frac{1}{\beta - \alpha}\left(\varphi(\Phi^{-1}(\beta)) - \varphi(\Phi^{-1}(\alpha))\right)\right)$$

for timepoint t, where φ and Φ denote the density and quantile function of the standard normal distribution, respectively. The second forecaster predicts $g_t = (g_{1,t}, g_{2,t}, g_{3,t})$,

Scoring function	$\phi'(x_3)$
S_1	$(\beta-\alpha)\tanh((\beta-\alpha)x_3)$
S_2	$(\beta-\alpha)(2/\pi)\arctan((\beta-\alpha)x_3)$
S_3	$(\beta-\alpha)(2\Phi((\beta-\alpha)x_3)-1)$
S_4	$(\beta-\alpha)(-\mathbb{1}\{x_3<c_1\}+\mathbb{1}\{x_3>c_2\}$ $+\mathbb{1}\{c_1\le x_3\le c_2\}2(x_3-(c_1+c_2)/2)/(c_2-c_1))$

Table 1: Examples of scoring functions. In all cases we choose $g_1(x_1)=x_1$ and $g_2(x_2)=x_2$. The parameters $c_1,c_2\in\mathbb{R}$ satisfy $c_1<c_2$.

where $g_{1,t}=f_{1,t}+\varepsilon_t$, $g_{2,t}=f_{2,t}+\varepsilon_t$ and $g_{3,t}=f_{3,t}+\varepsilon_t$ with $(\varepsilon_t)_{t=1,\dots,N}$ is independent normally distributed noise with mean zero and variance σ^2. The third forecaster, $h_t=(h_{1,t},h_{2,t},h_{3,t})$, bases his predictions on the unconditional distribution of Y_t, that is $\mathcal{N}(0,2)$. Therefore,

$$h_t=\left(\sqrt{2}\Phi^{-1}(\alpha),\sqrt{2}\Phi^{-1}(\beta),-\frac{\sqrt{2}}{\beta-\alpha}\left(\varphi(\Phi^{-1}(\beta))-\varphi(\Phi^{-1}(\alpha))\right)\right).$$

It is clear that the first forecaster dominates the second and the third forecaster, that is, it will be preferred under any consistent scoring function because in case of the first and the second forecaster, the first one is ideal with respect to the information set $\sigma(\mu_t,\varepsilon_t)$, whereas the second one is based on the same information set but is not ideal. In case of the first and the third forecaster, both forecasters are ideal but the information set of the first forecaster $\sigma(\mu_t)$ is larger than the one of the third forecaster which is the trivial σ-algebra (Holzmann and Eulert, 2014). It will depend on the size of the variance σ^2 whether the second or the third forecaster is preferred. Figures 1 and 2 provide Murphy diagrams of all forecasters computed from a sample of size $N=100'000$. They are in line with our theoretical considerations concerning the ranking of the three forecasts.

We compare predictive performance using Diebold-Mariano tests (Diebold and Mariano, 1995) based on the scoring functions in Table 1. We consider samples of size $N=250$ and repeat our experiment 10'000 times. In the left panel of Table 2, we consider the case that $\alpha=1-\beta=0.1$ where $\mathrm{RVaR}_{\alpha,\beta}$ is a trimmed mean. We report the ratio of rejections of the null hypothesis that forecaster i outperforms forecaster j, $i,j\in\{1,2,3\}$ at significance level 0.05. Analogously, in the right panel of Table 2, we consider the case that α,β are both close to zero, that is, $\alpha=0.01$ and $\beta=0.05$, which is a setting that is relevant if $\mathrm{RVaR}_{\alpha,\beta}$ is used as a risk measure. For the scoring function S_4, we have experimented a bit with the values c_1 and c_2 and report the results for the choices that worked best in our experiments. A systematic study on how to choose these two parameters goes beyond the scope of the present paper.

For the situation of the left panel of Table 2 concerning $\alpha=1-\beta=0.1$, we can see that forecaster 1 (2) outperforms forecaster 3 with a power of 1 (almost 1) for all

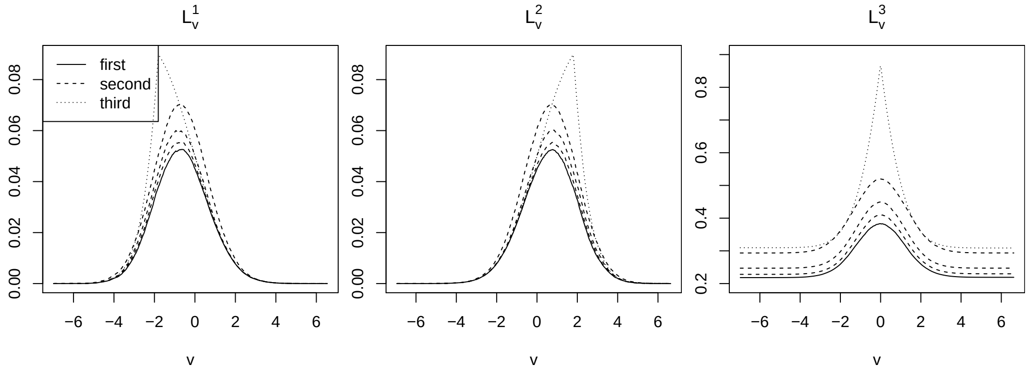

Figure 1: Murphy diagrams for $\alpha = 1 - \beta = 0.1$. Plots of expected elementary scores L_v^1, L_v^2, L_v^3 in terms of v for the three forecasters described in the test. For the second forecaster, the curves correspond to $\sigma = 0.3, 0.5, 0.8$ from bottom to top.

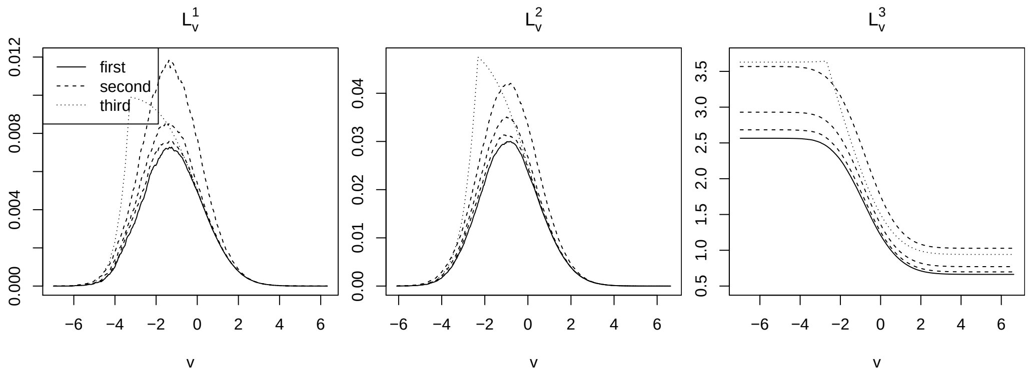

Figure 2: Murphy diagrams for $\alpha = 0.01$, $\beta = 0.05$. Plots of expected elementary scores L_v^1, L_v^2, L_v^3 in terms of v for the three forecasters described in the test. For the second forecaster, the curves correspond to $\sigma = 0.3, 0.5, 0.8$ from bottom to top.

H_0	S_1	S_2	S_3	S_4
$f \prec g$	0.304	0.406	0.417	0.624
$g \prec f$	0	0	0	0
$f \prec h$	1.000	1.000	1.000	1.000
$h \prec f$	0	0	0	0
$g \prec h$	0.999	0.998	0.992	0.998
$h \prec g$	0	0	0	0

H_0	S_1	S_2	S_3	S_4
$f \prec g$	0.515	0.529	0.500	0.566
$g \prec f$	0	0	0	0.003
$f \prec h$	0.995	1.000	0.996	0.835
$h \prec f$	0	0	0	0
$g \prec h$	0.874	0.993	0.885	0.393
$h \prec g$	0.001	0	0	0

Table 2: Power of Diebold-Mariano tests at significance level 0.05 for the scoring functions in Table 1 in the case that $\alpha = 1 - \beta = 0.1$ (left panel), and $\alpha = 0.01$, $\beta = 0.05$ (right panel). In the first case we chose $-c_1 = c_2 = 12$ for the scoring function S_4, and $c_1 = -5$, $c_2 = 1$ in the second case. The null hypothesis $f \prec g$ means that $\mathbb{E}S(f_t, Y_t) \le \mathbb{E}S(g_t, Y_t)$ for the scoring function specified in the column label. We chose $\sigma^2 = 0.5^2$ for the forecaster g.

scoring functions used. For a comparison of forecaster 1 and forecaster 2, the situation is more interesting: While forecaster 1 outperforms forecaster 2 with regard to all scoring functions considered, the power of the tests (and the associated discrimination ability of the scoring functions) varies substantially. While S_1 leads to an empirical power of 0.304 for the null hypothesis that $\mathbb{E}[S(f_t, Y_t)] \le \mathbb{E}[S(g_t, Y_t)]$, the score S_4 induces a power of 0.624 for the same null hypothesis.
The situation described in the right panel of Table 2 considering the parameter choice $\alpha = 0.01$ and $\beta = 0.05$ leads to a different situation. The tests employing S_1, S_2 and S_3 have a similar power. In contrast, S_4 yields a considerably smaller power (0.393) for the null $\mathbb{E}[S(h_t, Y_t)] \le \mathbb{E}[S(g_t, Y_t)]$ than the other scores (≥ 0.874 for all cases). A more detailed study and comparison of other scoring functions and other situations is deferred to future work.

7 Discussion

In this section, we would like to outline how one can implement our results about the elicitability of the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$, $0 < \alpha < \beta < 1$ in a regression context. Then we would like to contrast our ansatz to other suggestions for regression of the α-trimmed mean (which can be generalised to $\mathrm{RVaR}_{\alpha,\beta}$). The most common alternative approaches in the literature on robust statistics are the *trimmed least squares approach* and a two-step estimation procedure using the *Huber skipped mean*.

7.1 A joint regression framework for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$

Let $(X_t, Y_t)_{t\in\mathbb{N}}$ be a time series with the usual notation that Y_t denotes some real valued response variable and X_t is a d-dimensional vector of regressors. Let $\Theta \subseteq \mathbb{R}^k$ be some parameter space and $M\colon \mathbb{R}^d \times \Theta \to \mathsf{A}_0 = \{x \in \mathbb{R}^3 \,|\, x_1 \le x_3 \le x_2\}$ a parametric model. Let $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$, $0 < \alpha < b < 1$, and assume that there is a unique $\theta_0 \in \Theta$ such that

$$T(\mathcal{L}(Y_t|X_t)) = M(X_t, \theta_0) \quad \mathbb{P}\text{-a.s. for all } t \in \mathbb{N}, \tag{7.1}$$

where $\mathcal{L}(Y_t|X_t)$ denotes the conditional distribution of Y_t given X_t. That means, $M(X_t, \theta_0)$ models jointly the conditional VaR_α, VaR_β and the conditional $\mathrm{RVaR}_{\alpha,\beta}$. Let S be a strictly consistent scoring function of the form (3.3) and assume the sequence $(X_t, Y_t)_{t\in\mathbb{N}}$ satisfies certain mixing assumptions; see White (2001, Corollary 3.48). Then one obtains under additional integrability conditions that, as $n \to \infty$,

$$\frac{1}{n}\sum_{t=1}^{n} S(M(X_t,\theta), Y_t) - \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\big[S(M(X_t,\theta), Y_t)\big] \to 0 \quad \mathbb{P}\text{-a.s.}$$

It is essentially this Law of Large Numbers result which allows for consistent parameter estimation with the empirical estimator $\widehat{\theta}_n = \arg\min_{\theta\in\Theta} n^{-1}\sum_{t=1}^{n} S(M(X_t,\theta), Y_t)$; see e.g. van der Vaart (1998), Huber and Ronchetti (2009) and Nolde and Ziegel (2017) for details.

In summary, we can see that the complication of this procedure is that one needs to model the components VaR_α, VaR_β, even if one is only interested in $\mathrm{RVaR}_{\alpha,\beta}$. The advantage is that one can substantially deviate from an i.i.d. assumption on the data generating process. One can deal with non-independent, though mixing, and non-stationary data. The only amount of stationarity one needs is specified through (7.1).

7.2 Trimmed least squares

Most proposals for M-estimation and regression for $\mathrm{RVaR}_{\alpha,\beta}$ in the field of robust statistics focus on the α-trimmed mean, $\alpha \in (0, 1/2)$, corresponding to $\mathrm{RVaR}_{\alpha,1-\alpha}$. But they can often be extended to the general case $0 < \alpha < \beta < 1$ in a straightforward way. When this is the case, we describe the procedure in this more general manner. A majority of the proposals in the literature are commonly referred to as a *trimmed least squares* (TLS) approach. However, strictly speaking, TLS actually subsumes different, though closely related estimation procedures.

The first one was coined by Koenker and Basset (1978)—cf. Ruppert and Carroll (1980)—and constitutes a two step M-estimator: In a first step, the α- and β-quantile are determined via usual M-estimation. Then, all values below the former and above the latter are omitted and $\mathrm{RVaR}_{\alpha,\beta}$ is computed with an ordinary least squares approach. One can also express this procedure using order-statistics. Using the notation from Subsection 7.1, an M-estimator for $\mathrm{RVaR}_{\alpha,\beta}$ is given by

$$\arg\min_{x\in\mathbb{R}} \frac{1}{n}\sum_{i=[n\alpha]}^{[n\beta]} (x - Y_{(i)})^2.$$

Here, $Y_{(1)} \leq \cdots \leq Y_{(n)}$ is the order-statistics of the sample $Y_1, \ldots, Y_n$. While this procedure seems to work for a simplistic regression model (ignoring the regressors X_t and only modelling the intercept part), it is not clear how to use it in a more realistic regression context, where one is actually interested in the *conditional* distribution of Y_t given X_t rather than the unconditional distribution of Y_t. Moreover, this procedure only seems to work in an i.i.d. setting.

A second approach is described, for example, in Rousseeuw (1984, 1985) and relies on order-statistics of the squared residuals. It only seems to work for the α-trimmed mean. To be more precise, and again using the notation from above, let $m\colon \mathbb{R}^d \times \Theta \to \mathbb{R}$ be a one-dimensional parametric model. Again, one assumes that there is a unique correctly specified model parameter $\theta_0 \in \Theta$ such that

$$\mathrm{RVaR}_{\alpha,1-\alpha}(\mathcal{L}(Y_t|X_t)) = m(X_t, \theta_0) \quad \mathbb{P}\text{-a.s. for all } t \in \mathbb{N}. \tag{7.2}$$

For each $\theta \in \Theta$, define the residual $\varepsilon_t(\theta) := Y_t - m(X_t, \theta)$ and the absolute residual $r_t(\theta) := |\varepsilon_t(\theta)|$. Define the order-statistics of the absolute residuals $0 \leq r_{(1)}(\theta) \leq \cdots \leq r_{(n)}(\theta)$ for a sample of size n. Then an estimator is defined via

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n} \sum_{i=1}^{[2n\alpha]} r^2_{(i)}(\theta).$$

While this procedure appears to be fairly similar to an ordinary least squares procedure with the respective computational advantages, one should recall that the trimming crucially depends on the choice of the parameter θ. That means even if the model m is linear in the parameter θ, one generally yields a non-convex objective function with several local minima. Interestingly, the trimming takes place only for residuals with large modulus. If the error distribution, that is, the distribution of the residuals, is symmetric, this procedure yields a consistent estimator for θ_0 in an i.i.d. setting. If one wants to relax the assumption on the error distribution and is interested in modelling $\mathrm{RVaR}_{\alpha,\beta}$ for general $0 < \alpha < \beta < 1$ in (7.2), one could come up with the following ad-hoc procedure: Consider the order-statistics of the *residuals* $\varepsilon_{(1)}(\theta) \leq \cdots \leq \varepsilon_{(n)}(\theta)$. Then define an estimator via

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n} \sum_{i=[n\alpha]}^{[n\beta]} \left|\varepsilon_{(i)}(\theta)\right|^2.$$

This procedure takes into account the asymmetric nature of trimming when dealing with $\beta \neq 1-\alpha$ or $\beta = 1-\alpha$ and an asymmetric error distribution. However, one can see that the procedure cannot deal with the situation when the data generating process $(X_t, Y_t)_{t\in\mathbb{N}}$ is not stationary. In particular, heteroscedasticity can lead to severe problems. We would like to point out that, at the cost of additionally modelling the α- and β-quantile, the procedure using our strictly consistent scoring functions for the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ described in Subsection 7.1 does not rely on the usage of order-statistics and can in general deal with heteroscedasticity. The only degree of 'stationarity' is required through (7.1). Especially stationarity is deemed a too strong assumption in the context of financial data; see Davis (2016).

Finally, we would like to remark that there are further procedures belonging to the field of TLS. For instance, Atkinson and Cheng (1999) propose an adaptive procedure where the trimming parameter is data driven; see also Cerioli et al. (2018). However, we see no apparent way how to use such procedures if one is interested in predefined trimming parameters α and β.

7.3 Connections to Huber loss and Huber skipped mean

In his seminal paper, Huber (1964) introduced the famous *Huber loss* $S(x, y) = \rho(x-y)$ where $\rho(t) = \frac{1}{2}t^2$ for $|t| \le k$ and $\rho(t) = k|t| - \frac{1}{2}k^2$ for $|t| > k$. Huber argues that the "the corresponding [M-]estimator is related to Winsorizing" (Huber, 1964, p. 79). What obtained significantly less attention—maybe due to the lack of convexity—is another loss function he considers on the same page of the paper which is defined as $S(x, y) = \rho(x-y)$ for $\rho(t) = \frac{1}{2}t^2$ for $|t| \le k$ and $\rho(t) = \frac{1}{2}k^2$ for $|t| > k$. He writes about it: "the corresponding [M-]estimator is a trimmed mean" (*ibidem*).

One could define an asymmetric version of the latter loss function by using $S_{k_1,k_2}(x, y) = \rho_{k_1,k_2}(x-y)$ with

$$\rho_{k_1,k_2}(t) = \begin{cases} \frac{1}{2}k_1^2 & t < k_1 \\ \frac{1}{2}t^2 & k_1 \le t < k_2 \\ \frac{1}{2}k_2^2 & t \ge k_2. \end{cases}$$

The corresponding first-order condition for a minimum of the expected score $\bar{S}_{k_1,k_2}(x, F)$ is

$$\begin{aligned} 0 &= \int_{\{k_1 \le x-y < k_2\}} x - y \, \mathrm{d}F(y) = \int_{(x-k_2, x-k_1]} x - y \, \mathrm{d}F(y) \\ &= x\big(F(x-k_1) - F(x-k_2)\big) - \int_{(x-k_2, x-k_1]} y \, \mathrm{d}F(y). \end{aligned}$$

Now a suggestion similar to Rousseeuw (1984, p. 876) is to consider this loss with $k_1 = \mathrm{VaR}_\beta(F)$ and $k_2 = \mathrm{VaR}_\alpha(F)$ stemming from some pre-estimate. However, one can see that the first order-condition is generally not solved by $\mathrm{RVaR}_{\alpha,\beta}(F)$. Again, if one is interested in M-estimation for the trimmed mean or, more generally, RVaR, one should use the scoring functions introduced at (3.3).

Appendix

We present a list of assumptions used in Section 3. For more details about their interpretations and implications, please see Fissler and Ziegel (2016) were they were originally introduced.

Assumption (V1). Let $\mathcal{F}$ be a convex class of distribution functions on $\mathbb{R}$ and assume that for every $x \in \mathrm{int}(\mathsf{A})$ there are $F_1, \dots, F_{k+1} \in \mathcal{F}$ such that

$$0 \in \mathrm{int}\left(\mathrm{conv}\left(\left\{\bar{V}(x, F_1), \dots, \bar{V}(x, F_{k+1})\right\}\right)\right) .$$

Note that if $V\colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}^k$ is a strict $\mathcal{F}$-identification function for $T\colon \mathcal{F} \to \mathsf{A}$ which satisfies Assumption (V1), then for each $x \in \mathrm{int}(\mathsf{A})$ there is an $F \in \mathcal{F}$ such that $T(F) = x$.

Assumption (V3). For every $F \in \mathcal{F}$, the function $\bar{V}(\cdot, F)$ is continuously differentiable.

Assumption (V4). Let assumption (V3) hold. For all $r \in \{1, \dots, k\}$ and for all $t \in \mathrm{int}(\mathsf{A}) \cap T(\mathcal{F})$ there are $F_1, F_2 \in T^{-1}(\{t\})$ such that

$$\partial_l \bar{V}_l(t, F_1) = \partial_l \bar{V}_l(t, F_2) \quad \forall l \in \{1, \dots, k\} \setminus \{r\}, \qquad \partial_r \bar{V}_r(t, F_1) \neq \partial_r \bar{V}_r(t, F_2).$$

Assumption (F1). For every $y \in \mathbb{R}$ there exists a sequence $(F_n)_{n \in \mathbb{N}}$ of distributions $F_n \in \mathcal{F}$ that converges weakly to the Dirac-measure δ_y such that the support of F_n is contained in a compact set K for all n.

Assumption (VS1). Suppose that the complement of the set

$$C := \{(x, y) \in \mathsf{A} \times \mathbb{R} \mid V(x, \cdot) \text{ and } S(x, \cdot) \text{ are continuous at the point } y\}$$

has $(k + d)$-dimensional Lebesgue measure zero.

Assumption (S2). For every $F \in \mathcal{F}$, the function $\bar{S}(\cdot, F)$ is continuously differentiable and the gradient is locally Lipschitz continuous. Furthermore, $\bar{S}(\cdot, F)$ is twice continuously differentiable at $t = T(F) \in \mathrm{int}(\mathsf{A})$.

Acknowledgement

We would like to thank Timo Dimitriadis and Anthony C. Atkinson for insightful discussions about the topic, and Ruodu Wang, Rafael Frongillo and Tilmann Gneiting for helpful suggestions which improved an earlier version of this paper.
Tobias Fissler gratefully acknowledges funding of his Chapman Fellowship by the Department of Mathematics at Imperial College London, and Johanna Ziegel is grateful for financial support from the Swiss National Science Foundation.

References

C. Acerbi and B. Székely. Backtesting Expected Shortfall. *Risk Magazine*, 2014.

C. Acerbi and B. Székely. General properties of backtestable statistics. *Preprint*, 2017. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2905109.

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

A. C. Atkinson and T.-C. Cheng. Computing least trimmed squares regression with the forward search. *Statist. Comput.*, 9(4):251–263, 1999.

Bank for International Settlements. *Consultative Document: Fundamental review of the trading book: Outstanding issues.* 2014.

S. Barendse. Interquantile expectation regression. Technical report, Erasmus University Rotterdam, 2017. URL https://sites.google.com/site/scbarendse/papers.

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

J. R. Brehmer. Elicitability and its application in risk management. Master's thesis, University of Mannheim, 2017. URL http://arxiv.org/abs/1707.09604.

A. Cerioli, M. Riani, A. C. Atkinson, and A. Corbellini. The power of monitoring: how to make the most of a contaminated multivariate sample. *Stat. Methods Appl.*, 27(4):559–587, 2018.

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

M. H. A. Davis. Verification of internal risk measure estimates. *Stat. Risk Model.*, 33(3–4):67–93, 2016.

F. Delbaen, F. Bellini, V. Bignozzi, and J. F. Ziegel. Risk Measures with the CxLS property. *Finance Stoch.*, 20(433):433–453, 2016.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *J. Bus. Econom. Statist.*, 13: 253–263, 1995.

D. L. Donoho and P. J. Huber. The notion of breakdown point. In P. J. Bickel, K. A. Doksum, and J. L. Hodges, editors, *A Festschrift for Erich L. Lehmann*, Wadsworth, Belmont, 1983.

W. Ehm, T. Gneiting, A. Jordan, and F. Krüger. Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(3):505–562, 2016.

P. Embrechts, G. Puccetti, L. Rüschendorf, R. Wang, and A. Beleraj. An Academic Response to Basel 3.5. *Risks*, 2(1):25–48, 2014.

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

P. Embrechts, H. Liu, T. Mao, and R. Wang. Quantile-based risk sharing with heterogeneous beliefs. *Math. Program.*, 2018a.

P. Embrechts, H. Liu, and R. Wang. Quantile-based risk sharing. *Oper. Res.*, 66(4):936–949, 2018b.

S. Emmer, M. Kratz, and D. Tasche. What is the best risk measure in practice? A comparison of standard risk measures. *Journal of Risk (The)*, 8:31–60, 2015.

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

T. Fissler. *On Higher Order Elicitability and Some Limit Theorems on the Poisson and Wiener Space.* PhD thesis, University of Bern, 2017. URL http://biblio.unibe.ch/download/eldiss/17fissler_t.pdf.

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

T. Fissler and J. F. Ziegel. Erratum: Higher Order Elicitability and Osband's Principle. *Preprint*, 2019a. URL https://arxiv.org/abs/1901.08826.

T. Fissler and J. F. Ziegel. Order-Sensitivity and Equivariance of Scoring Functions. *Electron. J. Stat.*, Accepted for publication, 2019b.

T. Fissler, J. F. Ziegel, and T. Gneiting. Expected shortfall is jointly elicitable with value-at-risk: implications for backtesting. *Risk Magazine*, pages 58–61, January 2016.

R. Frongillo and I. Kash. On elicitation complexity and conditional elicitation. *Preprint*, 2015. URL https://arxiv.org/abs/1506.07212.

T. Gneiting. Making and Evaluating Point Forecasts. *J. Amer. Statist. Assoc.*, 106:746–762, 2011.

T. Gneiting and A. Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Amer. Statist. Assoc.*, 102:359–378, 2007.

T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 69:243–268, 2007.

F. R. Hampel. A General Qualitative Definition of Robustness. *Ann. Math. Statist.*, 42(6): 1887–1896, 12 1971.

C. Heinrich. The mode functional is not elicitable. *Biometrika*, 101(1):245–251, 2014.

H. Holzmann and M. Eulert. The role of the information set for forecasting – with applications to risk management. *Ann. Appl. Stat.*, 8:79–83, 2014.

P. J. Huber. Robust Estimation of a Location Parameter. *Ann. Math. Statist.*, 35(1):73–101, 03 1964.

P. J. Huber and E. M. Ronchetti. *Robust Statistics*. John Wiley & Sons, Inc., Hoboken, New Jersey, second edition, 2009.

J. M. Keynes. The principal averages and the laws of error which lead to them. *J. Roy. Statist. Soc.*, 74(3):322–331, 1911.

R. Koenker. *Quantile Regression*. Cambridge University Press, Cambridge, 2005.

R. Koenker and G. Basset. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

S. Kou, X. Peng, and C. C. Heyde. External Risk Measures and Basel Accords. *Math. Oper. Res.*, 38:393–417, 2013.

V. Krätschmer, A. Schied, and H. Zähle. Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *J. Multivariate Anal.*, 103:35–47, 2012.

V. Krätschmer, A. Schied, and H. Zähle. Comparative and qualitative robustness for law-invariant risk measures. *Finance Stoch.*, 18:271–295, 2014.

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

N. Lambert. Elicitation and Evaluation of Statistical Forecasts. *Preprint*, 2013. URL https://web.stanford.edu/~nlambert/papers/elicitation_july2013.pdf.

N. Lambert, D. M. Pennock, and Y. Shoham. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138, Chicago, Il, USA, 2008. ACM.

A. H. Murphy and H. Daan. Forecast Evaluation. In A. H. Murphy and R. W. Katz, editors, *Probability, Statistics and Decision Making in the Atmospheric Sciences*, pages 379–437. Westview Press, Boulder, Colorado, 1985.

W. K. Newey and J. L. Powell. Asymmetric Least Squares Estimation and Testing. *Econometrica*, 55:819–847, 1987.

N. Nolde and J. F. Ziegel. Elicitability and backtesting: Perspectives for banking regulation. *Ann. Appl. Stat.*, 11(4):1833–1874, 12 2017.

K. H. Osband. *Providing Incentives for Better Cost Forecasting*. PhD thesis, University of California, Berkeley, 1985.

A. J. Patton. Comparing possibly misspecified forecasts. *J. Bus. Econ. Stat. (forthcoming)*, 2019. URL http://public.econ.duke.edu/~ap172/.

P. J. Rousseeuw. Least Median of Squares Regression. *J. Amer. Statist. Assoc.*, 79(388):871–880, 1984.

P. J. Rousseeuw. Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vince, and W. Wertz, editors, *Mathematical Statistics and Applications*, pages 283–297. Reidel Publishing Company, Dordrecht, 1985.

D. Ruppert and R. J. Carroll. Trimmed Least Squares Estimation in the Linear Model. *J. Amer. Statist. Assoc.*, 75(372):828–838, 1980.

I. Steinwart, C. Pasin, R. Williamson, and S. Zhang. Elicitation and Identification of Properties. *JMLR: Workshop and Conference Proceedings*, 35:1–45, 2014.

J. W. Tukey. A survey of sampling from contaminated distributions. In I. Olkin et al., editor, *Contributions to Probability and Statistics*, pages 448–485. Stanford Univ. Press, 1960.

A. W. van der Vaart. *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. ISBN 0-521-49603-9.

R. Wang and Y. Wei. Risk functionals with convex level sets. *Preprint*, 2018. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3292661.

S. Weber. Distribution-Invariant Risk Measures, Information, and Dynamic Consistency. *Math. Finance*, 16:419–441, 2006.

H. White. *Asymptotic Theory for Econometricians*. Academic Press, San Diego, 2001.

H. Zähle. A definition of qualitative robustness for general point estimators, and examples. *J. Multivariate Anal.*, 143:12 – 31, 2016.

J. F. Ziegel. Coherence and elicitability. *Math. Finance*, 26(4):901–918, 2016.