Research Article

Tobias Fissler* and Johanna F. Ziegel

On the elicitability of range value at risk

Abstract: The debate of which quantitative risk measure to choose in practice has mainly focused on the dichotomy between value at risk (VaR) and expected shortfall (ES). Range value at risk (RVaR) is a natural interpolation between VaR and ES, constituting a tradeoff between the sensitivity of ES and the robustness of VaR, turning it into a practically relevant risk measure on its own. Hence, there is a need to statistically assess, compare and rank the predictive performance of different RVaR models, tasks subsumed under the term "comparative backtesting" in finance. This is best done in terms of strictly consistent loss or scoring functions, i.e., functions which are minimized in expectation by the correct risk measure forecast. Much like ES, RVaR does not admit strictly consistent scoring functions, i.e., it is not elicitable. Mitigating this negative result, we show that a triplet of RVaR with two VaR-components is elicitable. We characterize all strictly consistent scoring functions for this triplet. 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 regression.

Keywords: Backtesting, consistency, expected shortfall, point forecasts, scoring functions, trimmed mean

MSC 2010: 62C99, 62G35, 62P05, 91G70

1 Introduction

In the field of quantitative risk management, the last one or two decades have seen a lively debate about which monetary risk measure [3] would be best in (regulatory) practice. The debate mainly focused on the dichotomy between value at risk (VaR_β) on the one hand and expected shortfall (ES_β) on the other hand, at some probability level $\beta \in (0, 1)$ (see Section 2 for definitions). Mirroring the historical joust between median and mean as centrality measures in classical statistics, VaR_β, basically a quantile, is esteemed for its robustness, while ES_β, a tail expectation, is deemed attractive due to its sensitivity and the fact that it satisfies the axioms of a coherent risk measure [3]. We refer the reader to [15, 17] for comprehensive academic discussions, and to [58] for a regulatory perspective in banking.

Cont, Deguest and Scandolo [8] considered the issue of statistical robustness of risk measure estimates in the sense of [30]. They showed that a risk measure cannot be both robust and coherent. As a compromise, they propose the risk measure "range value at risk", $\text{RVaR}_{\alpha,\beta}$ at probability levels $0 < \alpha < \beta < 1$. It is defined as the average of all VaR_γ with γ between α and β (see Section 2 for definitions). As limiting cases, one obtains $\text{RVaR}_{\beta,\beta} = \text{VaR}_\beta$ and $\text{RVaR}_{0,\beta} = \text{ES}_\beta$, which presents $\text{RVaR}_{\alpha,\beta}$ as a natural interpolation of VaR_β and ES_β. Quantifying its robustness in terms of the *breakdown point* and following the arguments provided in [33, p. 59], $\text{RVaR}_{\alpha,\beta}$ has a breakdown point of $\min\{\alpha, 1 - \beta\}$, placing it between the very robust VaR_β (with a breakdown

*Corresponding author: Tobias Fissler, Department of Finance, Accounting and Statistics, Institute for Statistics and Mathematics, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria, e-mail: tobias.fissler@wu.ac.at. https://orcid.org/0000-0002-6541-7347
Johanna F. Ziegel, Department of Mathematics and Statistics, Institute of Mathematical Statistics and Actuarial Science, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland, e-mail: johanna.ziegel@stat.unibe.ch.
https://orcid.org/0000-0002-5916-9746

point of $\min\{\beta, 1 - \beta\}$) and the entirely non-robust ES_β (breakdown point 0). This means it is a robust – and hence not coherent – risk measure, unless it degenerates to $RVaR_{0,\beta} = ES_\beta$ (or if $0 \le \alpha < \beta = 1$). Moreover, RVaR belongs to the wide class of distortion risk measures [52, 55]. For further contributions to robustness in the context of risk measures, we refer the reader to [16, 36–38, 56]. Since the influential article [8], RVaR has gained increasing attention in the risk management literature – see [13, 14] for extensive studies – as well as in econometrics [5] where RVaR sometimes has the alternative denomination *interquantile expectation*. For the symmetric case $\beta = 1 - \alpha > \frac{1}{2}$, $RVaR_{\alpha,1-\alpha}$ is known under the term α-trimmed mean in classical statistics and it constitutes an alternative to and interpolation of the mean and the median as centrality measures; see [40] for a recent study and a multivariate extension of the trimmed mean. It is closely connected to the α-Winsorized mean; see (2.3).

How is it possible to evaluate the predictive performance of point forecasts X_t for a statistical functional T, such as the mean, median or a risk measure, of the (conditional) distribution of a quantity of interest Y_t? It is commonly measured in terms of the *average realized score* $\frac{1}{n} \sum_{t=1}^{n} S(X_t, Y_t)$ for some loss or scoring function S, using the orientation *the smaller the better*. Consequently, the loss function S should be *strictly consistent* for T in that $T(F) = \arg\min_x \int S(x, y)\, dF(y)$: correct predictions are honored and encouraged in the long run, e.g., the squared loss $S(x, y) = (x - y)^2$ is consistent for the mean, and the absolute loss $S(x, y) = |x - y|$ is consistent for the median. If a functional admits a strictly consistent score, it is called *elicitable* [27, 39, 44]. By definition, elicitable functionals allow for M-estimation and have natural estimation paradigms in regression frameworks [11, Section 2] such as quantile regression [34, 35] or expectile regression [42]. Elicitability is crucial for meaningful forecast evaluation [18, 27, 41]. In the context of probabilistic forecasts with distributional forecasts F_t or density forecasts f_t, (strictly) consistent scoring functions are often referred to as *(strictly) proper rules* such as the log-score $S(f, y) = -\log f(y)$ (see [29]). In quantitative finance, and particularly in the debate about which risk measure is best in practice, elicitability has gained considerable attention [9, 17, 57]. Especially, the role of elicitability for backtesting purposes has been highly debated [1, 2, 27]. It has been clarified that elicitability is central for *comparative* backtesting [24, 43]. On the other hand, if one strives to *validate* forecasts, (strict) *identification functions* are crucial. Much like scoring functions, they are functions in the forecast and the observation, which, however, vanish in expectation at (and only at) the correct report. Thus, they can be used to check (conditional) calibration [26, 43].

Not all functionals are elicitable or identifiable. Osband [44] showed that an elicitable or identifiable functional necessarily has convex level sets (CxLS): If $T(F_0) = T(F_1) = t$ for two distributions F_0, F_1, then $T(F_\lambda) = t$ where $F_\lambda = (1 - \lambda)F_0 + \lambda F_1$, $\lambda \in (0, 1)$. Variance and ES generally do not have CxLS [27, 53], therefore failing to be elicitable and identifiable. The *revelation principle* [19, 27, 44] asserts that any bijection of an elicitable/identifiable functional is elicitable/identifiable. This implies that the pair (mean, variance) – being a bijection of the first two moments – is elicitable and identifiable despite the variance failing to be so. Similarly, Fissler and Ziegel [21] showed that the pair (VaR_β, ES_β) is elicitable and identifiable, with the structural difference that the revelation principle is not applicable in this instance. This is followed by the more general finding that the minimal expected score and its minimizer are always jointly elicitable [6, 25].

Recently, Wang and Wei [51, Theorem 5.3] showed that $RVaR_{\alpha,\beta}$, $0 < \alpha < \beta < 1$, similarly to ES_α, fails to have CxLS as a standalone measure, which rules out its elicitability and identifiability. In contrast, they observe that the identity

$$RVaR_{\alpha,\beta} = \frac{\beta\, ES_\beta - \alpha\, ES_\alpha}{\beta - \alpha}, \quad 0 < \alpha < \beta < 1, \tag{1.1}$$

which holds if ES_α and ES_β are finite, and the CxLS property of the pairs (VaR_α, ES_α), (VaR_β, ES_β) implies the CxLS property of the triplet $(VaR_\alpha, VaR_\beta, RVaR_{\alpha,\beta})$ (see [51, Example 4.6]). This raises the question whether this triplet is elicitable and identifiable or not. By invoking the elicitability and identifiability of (VaR_α, ES_α), identity (1.1) and the revelation principle establish the elicitability and identifiability of the quadruples $(VaR_\alpha, VaR_\beta, ES_\alpha, RVaR_{\alpha,\beta})$ and $(VaR_\alpha, VaR_\beta, ES_\beta, RVaR_{\alpha,\beta})$. This approach has already been used in the context of regression in [5].

Improving this result, we show that the triplet $(VaR_\alpha, VaR_\beta, RVaR_{\alpha,\beta})$ is elicitable (Theorem 3.3) and identifiable (Proposition 3.1) under weak regularity conditions. Practically, our results open the way to model validation, to meaningful forecast performance comparison, and in particular to comparative backtests, of

this triplet, as well as to a regression framework. Theoretically, they show that the elicitation complexity [25, 39] or elicitation order [21] of $\mathrm{RVaR}_{\alpha,\beta}$ is at most 3. Moreover, requiring only VaR-forecasts besides the RVaR-forecast is particularly advantageous in comparison to additionally requiring ES-forecasts since the triplet $(\mathrm{VaR}_\alpha(F), \mathrm{VaR}_\beta(F), \mathrm{RVaR}_{\alpha,\beta}(F))$, $0 < \alpha < \beta < 1$, exists and is finite for any distribution F, whereas $\mathrm{ES}_\alpha(F)$ and $\mathrm{ES}_\beta(F)$ are only finite if the (left) tail of the gains-and-loss distribution F is integrable. As $\mathrm{RVaR}_{\alpha,\beta}$ is used often for robustness purposes, safeguarding against outliers and heavy-tailedness, this advantage is 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 [21] as well as the more general results of [6, 25]. 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 negative of a scaled *difference* of minima of expected strictly consistent scoring functions for VaR_α and VaR_β; see equations (3.1) and (3.2). 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. In particular, there is essentially no translation invariant or positively homogeneous scoring function which is strictly consistent for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$; see Section 4.

The paper is organized as follows. In Section 2, we introduce the relevant notation and definitions concerning RVaR, scoring functions and elicitability. The main results are presented in Section 3, establishing the elicitability of the triplet $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ (Theorem 3.3) and characterizing the class of strictly consistent scoring functions (Theorem 3.7), exploiting the identifiability result of Proposition 3.1. 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 [12]. 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 [35, 47, 49].

2 Notation and definitions

2.1 Definition of range value at risk

There are different sign conventions in the literature on risk measures. In this paper, we use the following convention: if a random variable Y models the gains and losses, then positive values of Y represent gains and negative values of Y losses. 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. In the sequel, let $\mathcal{F}_0$ be the class of probability distribution functions on $\mathbb{R}$. Recall that the α-quantile, $\alpha \in [0, 1]$, of $F \in \mathcal{F}_0$ is defined as the set $q_\alpha(F) = \{x \in \mathbb{R} \mid F(x-) \leq \alpha \leq F(x)\}$, where $F(x-) := \lim_{t \uparrow x} F(t)$.

Definition 2.1. *Value at risk* of $F \in \mathcal{F}_0$ at level $\alpha \in [0, 1]$ is defined by $\mathrm{VaR}_\alpha(F) = \inf q_\alpha(F)$.

For any $\alpha \in [0, 1]$ we introduce the following subclasses of $\mathcal{F}_0$:

$$\mathcal{F}^\alpha = \{F \in \mathcal{F}_0 \mid q_\alpha(F) = \{\mathrm{VaR}_\alpha(F)\}\}, \quad \mathcal{F}^{(\alpha)} = \{F \in \mathcal{F}_0 \mid F(\mathrm{VaR}_\alpha(F)) = \alpha\}.$$

Distributions $F \in \mathcal{F}^{(\alpha)}$ have at least one solution to the equation $F(x) = \alpha$; distributions $F \in \mathcal{F}^\alpha$ have at most one solution to the equation $F(x) = \alpha$.

Definition 2.2. *Range value at risk* of $F \in \mathcal{F}_0$ at levels $0 \leq \alpha \leq \beta \leq 1$ is defined by

$$\mathrm{RVaR}_{\alpha,\beta}(F) = \begin{cases} \dfrac{1}{\beta - \alpha} \displaystyle\int_{\alpha}^{\beta} \mathrm{VaR}_{\gamma}(F)\,\mathrm{d}\gamma & \text{if } \alpha < \beta, \\[4mm] \mathrm{VaR}_{\beta}(F) & \text{if } \alpha = \beta. \end{cases}$$

Note that $\lim_{\alpha \uparrow \beta} \mathrm{RVaR}_{\alpha,\beta}(F) = \mathrm{VaR}_{\beta}(F) = \mathrm{RVaR}_{\beta,\beta}(F)$. The definition of RVaR and the fact that $\gamma \mapsto \mathrm{VaR}_{\gamma}(F)$ is increasing imply that

$$\mathrm{VaR}_{\alpha}(F) \leq \mathrm{RVaR}_{\alpha,\beta}(F) \leq \mathrm{VaR}_{\beta}(F). \tag{2.1}$$

For $0 < \alpha \leq \beta < 1$ and $F \in \mathcal{F}_0$ one obtains that (i) $\mathrm{RVaR}_{\alpha,\beta}(F) \in \mathbb{R}$; (ii) $\mathrm{RVaR}_{0,\beta}(F) \in \mathbb{R} \cup \{-\infty\}$ and it is finite if and only if $\int_{-\infty}^{0} |y|\,\mathrm{d}F(y) < \infty$; and (iii) $\mathrm{RVaR}_{\alpha,1}(F) \in \mathbb{R} \cup \{\infty\}$ and it is finite if and only if $\int_{0}^{\infty} |y|\,\mathrm{d}F(y) < \infty$. Moreover, $\mathrm{RVaR}_{0,1}(F)$ exists if and only if

$$\int_{-\infty}^{0} |y|\,\mathrm{d}F(y) < \infty \quad \text{or} \quad \int_{0}^{\infty} |y|\,\mathrm{d}F(y) < \infty$$

and then coincides with $\int y\,\mathrm{d}F(y) \in \mathbb{R} \cup \{\pm\infty\}$. For $\alpha < \beta$ and provided that $\mathrm{RVaR}_{\alpha,\beta}(F)$ exists, it holds that

$$\mathrm{RVaR}_{\alpha,\beta}(F) = \frac{1}{\beta - \alpha} \left(\int\limits_{(\mathrm{VaR}_{\alpha}(F),\mathrm{VaR}_{\beta}(F)]} y\,\mathrm{d}F(y) + \mathrm{VaR}_{\alpha}(F)\big(F(\mathrm{VaR}_{\alpha}(F)) - \alpha\big) - \mathrm{VaR}_{\beta}(F)\big(F(\mathrm{VaR}_{\beta}(F)) - \beta\big) \right), \tag{2.2}$$

using the usual conventions $F(-\infty) = 0$, $F(\infty) = 1$ and $0 \cdot \infty = 0 \cdot (-\infty) = 0$. If $F \in \mathcal{F}^{(\alpha)} \cap \mathcal{F}^{(\beta)}$, then the correction terms in the second line of (2.2) vanish, yielding

$$\mathrm{RVaR}_{\alpha,\beta}(F) = \frac{\mathbb{E}_F(Y \mathbb{1}\{\mathrm{VaR}_{\alpha}(F) < Y \leq \mathrm{VaR}_{\beta}(F)\})}{\beta - \alpha},$$

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

Definition 2.3. *Expected shortfall* of $F \in \mathcal{F}_0$ at level $\alpha \in (0, 1)$ is defined by

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

Hence, provided that $\mathrm{ES}_{\alpha}(F)$ and $\mathrm{ES}_{\beta}(F)$ are finite, one obtains identity (1.1). 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 (1.1) 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 (1.1) is not defined.

Interestingly, [14, Theorem 2] establish that RVaR can be written as an inf-convolution of VaR and ES at appropriate levels. This result amounts to a sup-convolution in our sign convention. Also note that our parametrization of $\mathrm{RVaR}_{\alpha,\beta}$ differs from theirs.

Now, for $\alpha \in (0, \frac{1}{2})$, $\mathrm{RVaR}_{\alpha,1-\alpha}$ corresponds to the *α-trimmed mean* and has a close connection to the *α-Winsorized mean* W_{α} (see [33, pp. 57–59]) via

$$W_{\alpha}(F) := (1 - 2\alpha)\,\mathrm{RVaR}_{\alpha,1-\alpha}(F) + \alpha\,\mathrm{VaR}_{\alpha}(F) + \alpha\,\mathrm{VaR}_{1-\alpha}(F), \quad \alpha \in \left(0, \frac{1}{2}\right). \tag{2.3}$$

2.2 Elicitability and scoring functions

Using the decision-theoretic framework of [21, 27], we introduce the following notation. Let $\mathcal{F} \subseteq \mathcal{F}_0$ be some generic subclass and let $\mathsf{A} \subseteq \mathbb{R}^k$ be an *action domain*. 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. Then $T(\mathcal{F})$ corresponds to the image $\{T(F) \in \mathsf{A} \mid F \in \mathcal{F}\}$. 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 m-th partial derivative of $g(\cdot, y)$ by $\partial_m g(\cdot, y)$.

Definition 2.4. A map $S\colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ is an $\mathcal{F}$-*consistent scoring function* for $T\colon \mathcal{F} \to \mathsf{A}$ if it is $\mathcal{F}$-integrable and if $\bar{S}(T(F), F) \leq \bar{S}(x, F)$ for all $x \in \mathsf{A}$ and $F \in \mathcal{F}$. 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 $F \in \mathcal{F}$. A functional $T\colon \mathcal{F} \to \mathsf{A}$ is *elicitable* on $\mathcal{F}$ if it possesses a strictly $\mathcal{F}$-consistent scoring function.

Definition 2.5. Two scoring functions $S, \widetilde{S}\colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ are *equivalent* if there is some $a\colon \mathbb{R} \to \mathbb{R}$ and some $\lambda > 0$ such that $\widetilde{S}(x, y) = \lambda S(x, y) + a(y)$ for all $(x, y) \in \mathsf{A} \times \mathbb{R}$. They are *proportional* if they are equivalent with $a \equiv 0$.

This equivalence relation preserves (strict) consistency: If S is (strictly) $\mathcal{F}$-consistent for T and if a is $\mathcal{F}$-integrable, then $\widetilde{S}$ is also (strictly) $\mathcal{F}$-consistent for T. Closely related to the concept of elicitability is the notion of *identifiability*.

Definition 2.6. A map $V\colon \mathsf{A} \times \mathbb{R} \to \mathbb{R}^k$ is an $\mathcal{F}$-*identification function* for $T\colon \mathcal{F} \to \mathsf{A}$ if it is $\mathcal{F}$-integrable and if $\bar{V}(T(F), F) = 0$ for all $F \in \mathcal{F}$. It is a *strict* $\mathcal{F}$-identification function for T if additionally $\bar{V}(x, F) = 0$ implies that $x = T(F)$ for all $x \in \mathsf{A}$ and for all $F \in \mathcal{F}$. A functional $T\colon \mathcal{F} \to \mathsf{A}$ is *identifiable* on $\mathcal{F}$ if it possesses a strict $\mathcal{F}$-identification function.

In contrast to [27], we consider point-valued functionals only. For a recent comprehensive study on elicitability of set-valued functionals, we refer to [20].

3 Elicitability and identifiability results

Wang and Wei [51, Theorem 5.3] showed that for $0 < \alpha < \beta < 1$, $\mathrm{RVaR}_{\alpha,\beta}$ (and also the pairs $(\mathrm{VaR}_\alpha, \mathrm{RVaR}_{\alpha,\beta})$ and $(\mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$) do not have CxLS on $\mathcal{F}_{\mathrm{dis}}$, the class of distributions with bounded and discrete support. Hence, by invoking that CxLS are necessary both for elicitability and for identifiability, $\mathrm{RVaR}_{\alpha,\beta}$ and the pairs $(\mathrm{VaR}_\alpha, \mathrm{RVaR}_{\alpha,\beta})$ and $(\mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ are neither elicitable nor identifiable on $\mathcal{F}_{\mathrm{dis}}$. Our novel contribution is that the *triplet* $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$, however, is elicitable and identifiable, subject to mild conditions. We use the notation $S_\alpha(x, y) = (\mathbb{1}\{y \leq x\} - \alpha)x - \mathbb{1}\{y \leq x\}y$ and recall that S_α is $\mathcal{F}$-consistent for VaR_α if $\int_{-\infty}^0 |y| \, \mathrm{d}F(y) < \infty$ for all $F \in \mathcal{F}$, and strictly $\mathcal{F}$-consistent if furthermore $\mathcal{F} \subseteq \mathcal{F}^\alpha$ (see [27]).

Proposition 3.1. *For $0 < \alpha < \beta < 1$, the map $V\colon \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3$ defined by*

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

is an $\mathcal{F}^{(\alpha)} \cap \mathcal{F}^{(\beta)}$-identification function for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$, which is strict on $\mathcal{F}^\alpha \cap \mathcal{F}^{(\alpha)} \cap \mathcal{F}^\beta \cap \mathcal{F}^{(\beta)}$.

Proof. The proof is standard, observing that

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

which follows from the representation (2.2). □

Remark 3.2. The benefits of the identifiability result of Proposition 3.1 are two-fold. First, it facilitates (conditional) calibration backtests in the spirit of [43]. There, the null hypothesis is that a sequence of forecasts

$(X_{1,t}, X_{2,t}, X_{3,t})$, measurable with respect to the most recent information $\mathcal{A}_{t-1}$, is correctly specified in the sense that

$$(X_{1,t}, X_{2,t}, X_{3,t}) = (\mathrm{VaR}_\alpha(Y_t|\mathcal{A}_{t-1}), \mathrm{VaR}_\beta(Y_t|\mathcal{A}_{t-1}), \mathrm{RVaR}_{\alpha,\beta}(Y_t|\mathcal{A}_{t-1})).$$

By exploiting the strict identification property of V in (3.1), this null hypothesis corresponds to

$$\mathbb{E}\big(V(X_{1,t}, X_{2,t}, X_{3,t}, Y_t)|\mathcal{A}_{t-1}\big) = 0.$$

Clearly, such a conditional backtest can be conducted using any strict identification function. By invoking [19, Proposition 3.2.1], any strict $\mathcal{F}^\alpha \cap \mathcal{F}^{(\alpha)} \cap \mathcal{F}^\beta \cap \mathcal{F}^{(\beta)}$-identification function for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ is given by

$$H(x_1, x_2, x_3)V(x_1, x_2, x_3, y),$$

where V is given in (3.1) and $H\colon \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ is a matrix-valued function whose determinant does not vanish.

Second, Proposition 3.1 enables the characterization result of strictly consistent scoring functions presented in Theorem 3.7.

The following theorem establishes a rich class of (strictly) consistent scoring functions $S\colon \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$ for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$. By a priori assuming the forecasts to be bounded with values in some cube $[c_{\min}, c_{\max}]^3$, $-\infty \le c_{\min} < c_{\max} \le \infty$ (here and throughout the paper, we make the tacit convention that $[c_{\min}, c_{\max}] := [c_{\min}, c_{\max}] \cap \mathbb{R}$ if $c_{\min} = -\infty$ or $c_{\max} = \infty$), the class gets even broader.

Theorem 3.3. *For $0 < \alpha < \beta < 1$, the map $S\colon [c_{\min}, c_{\max}]^3 \times \mathbb{R} \to \mathbb{R}$ defined by*

$$S(x_1, x_2, x_3, y) = (\mathbb{1}\{y \le x_1\} - \alpha)g_1(x_1) - \mathbb{1}\{y \le x_1\}g_1(y) + (\mathbb{1}\{y \le x_2\} - \beta)g_2(x_2) - \mathbb{1}\{y \le x_2\}g_2(y)$$

$$+ \phi'(x_3)\Big(x_3 + \frac{1}{\beta - \alpha}(S_\beta(x_2, y) - S_\alpha(x_1, y))\Big) - \phi(x_3) + a(y) \tag{3.3}$$

is an $\mathcal{F}$-consistent scoring function for $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ if the following conditions hold:
(i) $\phi\colon [c_{\min}, c_{\max}] \to \mathbb{R}$ is convex with subgradient ϕ'.
(ii) For all $x_3 \in [c_{\min}, c_{\max}]$ the functions

$$G_{1,x_3}\colon [c_{\min}, c_{\max}] \to \mathbb{R}, \quad x_1 \mapsto g_1(x_1) - x_1\phi'(x_3)/(\beta - \alpha), \tag{3.4}$$

$$G_{2,x_3}\colon [c_{\min}, c_{\max}] \to \mathbb{R}, \quad x_2 \mapsto g_2(x_2) + x_2\phi'(x_3)/(\beta - \alpha), \tag{3.5}$$

are increasing.
(iii) $y \mapsto a(y) - \mathbb{1}\{y \le x_1\}g_1(y) - \mathbb{1}\{y \le x_2\}g_2(y)$ is $\mathcal{F}$-integrable for all $x_1, x_2 \in [c_{\min}, c_{\max}]$.
If moreover ϕ is strictly convex and the functions in G_{1,x_3} and G_{2,x_3} in (3.4) and (3.5) are strictly increasing for all $x_3 \in [c_{\min}, c_{\max}]$, then S is strictly $\mathcal{F}^\alpha \cap \mathcal{F}^\beta$-consistent for T.

Proof. 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,

$$[c_{\min}, c_{\max}] \times \mathbb{R} \ni (x_1', y) \mapsto S(x_1', x_2, x_3, y)$$

is $\mathcal{F}$-consistent for VaR_α and it is strictly $\mathcal{F}^\alpha$-consistent if G_{1,x_3} is strictly increasing. Similar comments apply to the map $[c_{\min}, c_{\max}] \times \mathbb{R} \ni (x_2', y) \mapsto S(t_1, x_2', x_3, y)$. Hence,

$$0 \le \bar{S}(x_1, x_2, x_3, F) - \bar{S}(t_1, x_2, x_3, F) + \bar{S}(t_1, x_2, x_3, F) - \bar{S}(t_1, t_2, x_3, F)$$

$$= \bar{S}(x_1, x_2, x_3, F) - \bar{S}(t_1, t_2, x_3, F),$$

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 $x_3 \ne t_3$, the inequality in (3.6) is strict. □

Remark 3.4. Provided condition (iii) in Theorem 3.3 holds and if ϕ is strictly convex and G_{1,x_3} and G_{2,x_3} are strictly increasing, then S given in (3.3) is still strictly $\mathcal{F}$-consistent in the RVaR-component for general

$\mathcal{F} \subseteq \mathcal{F}_0$. That is, for $F \in \mathcal{F}$,

$$\underset{x \in A_0}{\arg\min}\, \bar{S}(x, F) = q_\alpha(F) \times q_\beta(F) \times \{\mathrm{RVaR}_{\alpha,\beta}(F)\}.$$

By making use of (2.3) and the revelation principle [19, 27, 44], Theorem 3.3 also provides a rich class of strictly consistent scoring functions for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_{1-\alpha}, W_\alpha)$, where W_α is the α-Winsorized mean. The following proposition is useful to construct examples; see Section 6.

Proposition 3.5. *Let S be of the form* (3.3) *with a (strictly) convex and non-constant function ϕ, and functions g_1, g_2 such that the functions at* (3.4) *and* (3.5) *are (strictly) increasing and condition* (iii) *of Theorem 3.3 is satisfied. Then the following assertions hold:*

(i) *The subgradient ϕ' of ϕ is necessarily bounded and the one-sided derivatives of g_1 and g_2 are necessarily bounded from below.*

(ii) *S is proportional 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*

$$\beta - \alpha = -\inf_{x \in [c_{\min}, c_{\max}]} \tilde{\phi}'(x) = \sup_{x \in [c_{\min}, c_{\max}]} \tilde{\phi}'(x),$$

and strictly increasing functions $\tilde{g}_1, \tilde{g}_2$ such that their one-sided derivatives are bounded from below by one and such that the functions at (3.4) *and* (3.5) *are (strictly) increasing and condition* (iii) *of Theorem 3.3 is satisfied.*

Proof. (i) The proof is similar to the one of [21, Corollary 5.5]: condition (ii) implies that for any

$$x_1, x_1', x_2, x_2', x_3 \in [c_{\min}, c_{\max}]$$

with $x_1 < x_1'$ and $x_2 < x_2'$ it holds that

$$-\infty < -\frac{g_2(x_2') - g_1(x_2)}{x_2' - x_2} \le \frac{\phi'(x_3)}{\beta - \alpha} \le \frac{g_1(x_1') - g_1(x_1)}{x_1' - x_1} < \infty.$$

Therefore, ϕ' is bounded, and the one-sided derivative of g_1 is bounded from below by $\sup_{x_3} \phi'(x_3)/(\beta - \alpha)$, while the one-sided derivative of g_2 is bounded from below by $-\inf_{x_3} \phi'(x_3)/(\beta - \alpha)$.

(ii) For any $c \in \mathbb{R}$, if we replace ϕ by $\widehat{\phi} : x \mapsto \phi(x) + cx$, g_1 by $\widehat{g}_1 : x \mapsto g_1(x) + cx/(\beta - \alpha)$, and g_2 by $\widehat{g}_2 : x \mapsto g_2(x) - cx/(\beta - \alpha)$ in formula (3.3) for S, then S does not change. Also, $\widehat{\phi}$ is (strictly) convex if and only if ϕ is (strictly) convex. Furthermore, conditions (ii) and (iii) of Theorem 3.3 hold for (ϕ, g_1, g_2) if and only if they hold for $(\widehat{\phi}, \widehat{g}_1, \widehat{g}_2)$. By part (i) of the proposition, ϕ' is bounded. Therefore, we can assume without loss of generality that

$$-\inf_{x \in [c_{\min}, c_{\max}]} \phi'(x) = \sup_{x \in [c_{\min}, c_{\max}]} \phi'(x) = \lambda > 0$$

since ϕ is non-constant. Then the argument follows by setting $\tilde{S} = \frac{\lambda}{\beta - \alpha} S$. □

Example 3.6. Proposition 3.5 in combination with Theorem 3.3 yields a straightforward recipe to generate (strictly) consistent scoring functions for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$. The main degree of flexibility is the choice of ϕ. For practical purposes, it can be easier to start with the choice of ϕ', which should be a (strictly) increasing and bounded function. A rich source for such functions is the class of (strictly increasing) cumulative distribution functions, which can easily be scaled to have an infimum of $-(\beta - \alpha)$ and a supremum of $\beta - \alpha$. Then ϕ can be obtained by integrating ϕ'. The simplest choice for g_1 and g_2 is the identity, i.e., $g_1(x_1) = x_1$ and $g_2(x_2) = x_2$. The only remaining degree of flexibility is then to add consistent scoring functions for VaR_α or for VaR_β. Table 1 contains some examples for choices of ϕ'. For illustrative purposes, let us discuss the score S_1 from Table 1 more closely. Just as in the case of S_3, but less obviously so, the corresponding ϕ' is motivated by a distribution function. In this case, it is the logistic distribution $e^{x_3}/(1 + e^{x_3})$. Proper translation and scaling according to Proposition 3.5 leads to

$$\phi'(x_3) = (\beta - \alpha)\Big(\frac{2e^{x_3}}{1 + e^{x_3}} - 1\Big) = (\beta - \alpha)\frac{e^{x_3} - 1}{e^{x_3} + 1} = (\beta - \alpha)\tanh\Big(\frac{x_3}{2}\Big).$$

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

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 $c_1 < c_2$, and Φ is the cumulative distribution function of a standard normal law.

An antiderivative of ϕ' is given by $\phi(x_3) = (\beta - \alpha)(2\log(e^{x_3} + 1) - x_3)$. Therefore, upon choosing $a(y) = 2y$, the explicit form of S_1 reads

$$S_1(x_1, x_2, x_3, y) = (\mathbb{1}\{y \leq x_1\} - \alpha)x_1 + \mathbb{1}\{y > x_1\}y + (\mathbb{1}\{y \leq x_2\} - \beta)x_2$$

$$+ \mathbb{1}\{y > x_2\}y + 2(\beta - \alpha)\Big(x_3\frac{e^{x_3}}{e^{x_3} + 1} - \log(e^{x_3} + 1)\Big)$$

$$+ \frac{e^{x_3} - 1}{e^{x_3} + 1}((\mathbb{1}\{y \leq x_2\} - \beta)x_2 - \mathbb{1}\{y \leq x_2\}y - (\mathbb{1}\{y \leq x_1\} - \alpha)x_1 + \mathbb{1}\{y \leq x_1\}y).$$

The particular choice of $a(y) = 2y$ can be beneficial with regard to integrability conditions: With this choice, S_1 is F-integrable if and only if the right tail of F is integrable, i.e., if $\int_0^\infty y \, dF(y) < \infty$. In a risk management context with our sign convention, the right tail corresponds to the gains, which are commonly less heavy-tailed than the losses. While ϕ' appearing in S_2 can easily be integrated with an antiderivative of

$$\phi(x_3) = (\beta - \alpha)\Big(\frac{2}{\pi}\Big)\Big(x_3 \arctan(x_3) - \frac{\log(x_3^2 + 1)}{2}\Big),$$

the antiderivative of ϕ' for S_3 has no closed form solution, therefore requiring numerical integration. The scoring function S_4, where ϕ' is an increasing piecewise linear function which is strictly increasing only on $[c_1, c_2]$, is in the spirit of the Huber loss [32, p. 79]. It is only strictly consistent on $[c_1, c_2]^3$, but remains consistent for all of $\mathbb{R}^3$.

Striving for a full characterization of the class of strictly consistent scoring functions for

$$T = (\text{VaR}_\alpha, \text{VaR}_\beta, \text{RVaR}_{\alpha,\beta}),$$

we shall next establish the counterpart of Theorem 3.3, providing necessary conditions for the strict consistency. The main tool to derive such necessary conditions is *Osband's principle*, originating from the seminal dissertation of Osband [44]; see also [27] for an accessible intuition. We use the precise technical formulation of [21, Theorem 3.2]. It is no wonder that necessary conditions for strictly $\mathcal{F}$-consistent scores for T can only be obtained for action domains $\mathsf{A} \subseteq \mathbb{R}^3$ such that the surjectivity condition $\mathsf{A} = \{T(F): F \in \mathcal{F}\}$ holds. By invoking inequality (2.1), any such action domain is necessarily a subset of

$$\mathsf{A}_0 := \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 \leq x_3 \leq x_2\},$$

which we therefore call the *maximal sensible action domain*. Issuing forecasts for T outside of A_0, thus violating (2.1), would be irrational, comparable to, say, negative variance forecasts. Still, the scoring functions of the form (3.3) allow for the evaluation of forecasts violating (2.1). Besides the surjectivity assumption and further richness assumptions on the class of distributions $\mathcal{F}$, we need to impose smoothness conditions on the expected score as to exploit first-order conditions stemming from the minimization problem of strict consistency; see Section A for the detailed technical formulations and [21] for a discussion of these conditions.

We introduce the class $\mathcal{F}_{\text{cont}}^\alpha \subset \mathcal{F}^\alpha$ of distributions in $\mathcal{F}^\alpha$ which are continuously differentiable (and therefore also in $\mathcal{F}^{(\alpha)}$). For any $\mathsf{A} \subseteq \mathbb{R}^3$, we denote the projections on the r-th component by

$$\mathsf{A}_r' := \{x_r \in \mathbb{R} \mid \text{there exists } (z_1, z_2, z_3) \in \mathsf{A}, \ z_r = x_r\}, \quad r \in \{1, 2, 3\}.$$

For any $x_3 \in A_3'$ and $m \in \{1, 2\}$, let

$$A_{m,x_3}' := \{x_m \in \mathbb{R} \mid \text{there exists } (z_1, z_2, z_3) \in A, \, z_m = x_m, \, z_3 = x_3\}.$$

Theorem 3.7. *Let $\mathcal{F} \subseteq \mathcal{F}_{cont}^{\alpha}$, $0 < \alpha < \beta < 1$, $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta}) \colon \mathcal{F} \to A \subseteq A_0$, and let $V = (V_1, V_2, V_3)^{\mathsf{T}}$ be defined at (3.1). If Assumptions (V1) and (F1) hold and $(V_1, V_2)^{\mathsf{T}}$ satisfies Assumption (V4), then any strictly $\mathcal{F}$-consistent scoring function $S \colon A \times \mathbb{R} \to \mathbb{R}$ for T that satisfies Assumptions (VS1) and (S2) is necessarily of the form (3.3) almost everywhere, where the functions $G_{r,x_3} \colon A_{r,x_3}' \to \mathbb{R}$, $r \in \{1, 2\}$, $x_3 \in A_3'$, in (3.4) and (3.5) are strictly increasing and $\phi \colon A_3' \to \mathbb{R}$ is strictly convex.*

Proof. First note that V satisfies Assumption (V3) on $\mathcal{F} \subseteq \mathcal{F}_{cont}^{\alpha}$. Let $F \in \mathcal{F}$ with derivative f and let $x \in \mathrm{int}(A)$. Then one obtains

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

The partial derivatives of V are given by

$$\partial_1 \bar{V}_1(x, F) = f(x_1),$$

$$\partial_2 \bar{V}_2(x, F) = f(x_2),$$

$$\partial_1 \bar{V}_3(x, F) = -\frac{F(x_1) - \alpha}{\beta - \alpha},$$

$$\partial_2 \bar{V}_3(x, F) = \frac{F(x_2) - \beta}{\beta - \alpha},$$

$$\partial_3 \bar{V}_3(x, F) = 1,$$

and $\partial_r \bar{V}_1(x, F)$ and $\partial_m \bar{V}_2(x, F)$ vanish for $r \in \{2, 3\}$ and $m \in \{1, 3\}$. Applying [21, Theorem 3.2] yields the existence of continuously differentiable functions $h_{lm} \colon \mathrm{int}(A) \to \mathbb{R}$, $l, m \in \{1, 2, 3\}$, such that

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

for $m \in \{1, 2, 3\}$. 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 yields that $h_{12} \equiv h_{21} \equiv 0$ on $\mathrm{int}(A)$. 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

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

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

$$h_{13} \equiv h_{31} \equiv h_{23} \equiv h_{32} \equiv 0.$$

So we are left with characterizing h_{mm} for $m \in \{1, 2, 3\}$. Note that Assumption (V1) implies that for any $x = (x_1, x_2, x_3) \in \mathrm{int}(A)$ there are two distributions $F_1, F_2 \in \mathcal{F}$ such that

$$(F_1(x_1) - \alpha, F_1(x_2) - \beta)^{\mathsf{T}} \quad \text{and} \quad (F_2(x_1) - \alpha, F_2(x_2) - \beta)^{\mathsf{T}}$$

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

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

Again, Assumption (V1) implies that there are $F_1, F_2 \in \mathcal{F}$ such that

$$(\bar{V}_1(x, F_1), \bar{V}_3(x, F_1))^\mathsf{T} \quad \text{and} \quad (\bar{V}_1(x, F_2), \bar{V}_3(x, F_2))^\mathsf{T}$$

are linearly independent. Hence, we obtain that $\partial_1 h_{33} \equiv 0$ and $\partial_3 h_{11} \equiv -h_{33}/(\beta - \alpha)$. 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 $\partial_2 h_{33} \equiv 0$ and $\partial_3 h_{22} \equiv h_{33}/(\beta - \alpha)$. This means there exist functions

$$\eta_1 \colon \{(x_1, x_3) \in \mathbb{R}^2 \mid \text{there exists } (z_1, z_2, z_3) \in \text{int}(A),\ x_1 = z_1,\ x_3 = z_3\} \to \mathbb{R},$$
$$\eta_2 \colon \{(x_2, x_3) \in \mathbb{R}^2 \mid \text{there exists } (z_1, z_2, z_3) \in \text{int}(A),\ x_2 = z_2,\ x_3 = z_3\} \to \mathbb{R},$$
$$\eta_3 \colon \text{int}(A)_3' \to \mathbb{R},$$

and some $z \in \text{int}(A)_3'$ such that for any $x = (x_1, x_2, x_3) \in \text{int}(A)$ it holds that

$$h_{33}(x) = \eta_3(x_3),$$

$$h_{11}(x) = \eta_1(x_1, x_3) = -\frac{1}{\beta - \alpha} \int_z^{x_3} \eta_3(z)\, dz + \zeta_1(x_1),$$

$$h_{22}(x) = \eta_2(x_2, x_3) = \frac{1}{\beta - \alpha} \int_z^{x_3} \eta_3(z)\, dz + \zeta_2(x_2),$$

where $\zeta_r \colon \text{int}(A)_r' \to \mathbb{R}$, $r \in \{1, 2\}$. Due to the fact that any component of T is mixture-continuous[1] and since $\mathcal{F}$ is convex and T surjective, the projection $\text{int}(A)_3'$ is an open interval. Hence, $[\min(z, x_3), \max(z, x_3)] \subset \text{int}(A)_3'$. Due to Assumptions (V3) and (S2), [21, Theorem 3.2] implies that η_1, η_2, η_3 are locally Lipschitz continuous.

The above calculations imply that the Hessian of the expected score, i.e., $\nabla^2 \bar{S}(x, F)$, at its minimizer $x = t = T(F)$, is a diagonal matrix with entries $\eta_1(t_1, t_3) f(t_1)$, $\eta_2(t_2, t_3) f(t_2)$ and $\eta_3(t_3)$. As a second-order condition, $\nabla^2 \bar{S}(t, F)$ must be positive semi-definite. By invoking the surjectivity of T once again, this shows that $\eta_1, \eta_2, \eta_3 \geq 0$. 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 $\frac{d}{ds} \bar{S}(t + sv, F)$ is negative for all $s \in (-\varepsilon, 0)$, zero for $s = 0$, and positive for all $s \in (\varepsilon, 0)$. For $v = e_3 = (0, 0, 1)^\mathsf{T}$, this means that for any $F \in \mathcal{F}$ with $t = T(F)$ there is an $\varepsilon > 0$ such that $\frac{d}{ds} \bar{S}(t + se_3, F) = \eta_3(t_3 + s)s$ has the same sign as s for all $s \in (-\varepsilon, \varepsilon)$. Therefore, $\eta_3(t_3 + s) > 0$ for all $s \in (-\varepsilon, \varepsilon) \setminus \{0\}$. Using the surjectivity of T and invoking a compactness argument, η_3 attains a 0 only finitely many times on any compact interval. Recall that $\text{int}(A)_3'$ is an open interval. Hence, it can be approximated by an increasing sequence of compact intervals. Therefore, $\eta_3^{-1}(\{0\})$ is at most countable, and therefore a Lebesgue null set. With similar arguments, one can show that for any $x_3 \in \text{int}(A)_3'$ the sets

$$\{x_1 \in \mathbb{R} \mid \text{there exists } (z_1, z_2, z_3) \in \text{int}(A),\ x_1 = z_1,\ x_3 = z_3,\ \eta_1(x_1, x_3) = 0\},$$
$$\{x_2 \in [x_3, \infty) \mid \text{there exists } (z_1, z_2, z_3) \in \text{int}(A),\ x_2 = z_2,\ x_3 = z_3,\ \eta_2(x_2, x_3) = 0\}$$

are at most countable, and therefore also Lebesgue null sets.

Finally, using [23, Proposition 1 in the supplement] (recognizing that V is locally bounded), one obtains that S is almost everywhere of the form (3.3). Moreover, it holds almost everywhere that $\phi'' = \eta_3$ and $g_m' = \zeta_m$ for $m \in \{1, 2\}$. Hence, ϕ is strictly convex and the functions at (3.4) and (3.5) are strictly increasing. □

---

1 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.

Combining Theorems 3.3 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

$$\mathsf{A} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid c_{\min} \le x_1 \le x_3 \le x_2 \le c_{\max}\}.$$

Corollary 3.8. *Let*

$$\mathsf{A} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid c_{\min} \le x_1 \le x_3 \le x_2 \le c_{\max}\}$$

for some $-\infty \le c_{\min} < c_{\max} \le \infty$. Under the conditions of Theorem 3.7, 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 (3.3) almost everywhere satisfying conditions (i)–(iii). *Moreover, the function $\phi'\colon [c_{\min}, c_{\max}] \to \mathbb{R}$ is necessarily bounded.*

Proof. For the proof it suffices to show that for $r \in \{1, 2\}$, G_{r,x_3} defined in (3.4) and (3.5) is not only increasing on A'_{r,x_3} for any $x_3 \in \mathsf{A}'_3$, but on $\mathsf{A}'_r = [c_{\min}, c_{\max}]$. For $x_3 \in [c_{\min}, c_{\max}] = \mathsf{A}'_3$, we have $\mathsf{A}'_{1,x_3} = [c_{\min}, x_3]$ and $\mathsf{A}'_{2,x_3} = [x_3, c_{\max}]$. 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

$$0 \le g_1(x'_1) - g_1(x_1) - (x'_1 - x_1)\frac{\phi'(x'_1)}{\beta - \alpha}$$

$$\le g_1(x'_1) - g_1(x_1) - (x'_1 - x_1)\frac{\phi'(x_3)}{\beta - \alpha},$$

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. $\square$

Remark 3.9. Note the structural difference of Theorems 3.3 and 3.7 to [25, Theorem 1], [6, Proposition 4.14] and in particular [21, 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 an expected scoring function – or *Bayes risk* –, 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 function in [21, Theorem 5.2] may well be set to zero. [25, Theorem 2] generalizes our results and presents an elicitability result of any linear combination of Bayes risks.

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 imposing secondary desirable criteria on S, e.g., acknowledging that $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ is translation equivariant (meaning that $T(F_{Y+z}) = T(F_Y) + z$ for any constant $z \in \mathbb{R}$) and positively homogeneous of degree 1 (meaning that $T(F_{cY}) = cT(F_Y)$ for any $c > 0$), it would make sense if the forecast ranking were also invariant under a joint translation of the forecasts and the observations on the one hand, and joint scaling of the forecasts and the observations on the other hand. This would require translation *invariance* of the score differences on the one hand, i.e.,

$$S(x_1 + z, x_2 + z, x_3 + z, y + z) - S(x'_1 + z, x'_2 + z, x'_3 + z, y + z) = S(x_1, x_2, x_3, y) - S(x'_1, x'_2, x'_3, y)$$

for all $(x_1, x_2, x_3), (x'_1, x'_2, x'_3) \in \mathsf{A}$ and $y, z \in \mathbb{R}$. On the other hand, it would require positively homogeneous score differences, that is, there is some $b \in \mathbb{R}$ such that

$$S(cx_1, cx_2, cx_3, cy) - S(cx'_1, cx'_2, cx'_3, cy) = c^b(S(x_1, x_2, x_3, y) - S(x'_1, x'_2, x'_3, y))$$

for all $(x_1, x_2, x_3), (x'_1, x'_2, x'_3) \in A$, $y \in \mathbb{R}$ and for all $c > 0$. While translation invariance seems to be particularly important when RVaR is used as a location parameter, i.e., when $\alpha = 1 - \beta < \frac{1}{2}$, corresponding to the α-trimmed mean, positively homogeneous score differences are relevant in a risk management context: the forecast ranking should not depend on the unit in which the risk measures and the gains and losses are reported, be it in, say Euros or in Euro Cents. We also refer to [22, 43, 45] for further motivations. This section establishes that, unfortunately, there are no strictly consistent scoring functions for $(\text{VaR}_\alpha, \text{VaR}_\beta, \text{RVaR}_{\alpha,\beta})$ which admit translation invariant or positively homogeneous score differences under practically relevant settings.

If one is interested in scoring functions with an action domain of the form

$$A = \{x \in \mathbb{R}^3 \mid c_{\min} \le x_1 \le x_3 \le x_2 \le c_{\max}\}$$

possessing the additional property of translation invariant score differences, the only sensible choice is $c_{\min} = -\infty$, $c_{\max} = \infty$, 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

$$A = A_0,$$
$$A = A_0^+ = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid 0 \le x_1 \le x_3 \le x_2\},$$
$$A = A_0^- = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 \le x_3 \le x_2 \le 0\}.$$

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

Proof. By using Theorem 3.7, any strictly $\mathcal{F}$-consistent scoring function for the functional

$$T = (\text{VaR}_\alpha, \text{VaR}_\beta, \text{RVaR}_{\alpha,\beta})$$

must be of the form (3.3), where in particular ϕ is strictly convex, twice differentiable and ϕ' is bounded. Assume that S has translation invariant score differences. That means that the function

$$\Psi \colon \mathbb{R} \times A_0 \times A_0 \times \mathbb{R} \to \mathbb{R}$$

defined by

$$\Psi(z, x, x', y) = S(x_1 + z, x_2 + z, x_3 + z, y + z) - S(x'_1 + z, x'_2 + z, x'_3 + z, y + z)$$
$$- S(x_1, x_2, x_3, y) + S(x'_1, x'_2, x'_3, y)$$

vanishes. Then, for all $x \in A_0$ and for all $z, y \in \mathbb{R}$,

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

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

The proof of Proposition 4.1 closely follows the one of [22, Proposition 4.10]. The fact that the latter assertion entails a positive result has the following background: The strictly consistent scoring function for $(\text{VaR}_\alpha, \text{ES}_\alpha)$ given in [22, Proposition 4.10] works only on a very restricted action domain. To guarantee strict consistency on such an action domain, one would need a refinement of Theorem 3.3 in the spirit of [23, Proposition 2 of the supplement]. 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 (Positive homogeneity). *Under the conditions of Theorem 3.7 there are no strictly $\mathcal{F}$-consistent scoring functions for $(\text{VaR}_\alpha, \text{VaR}_\beta, \text{RVaR}_{\alpha,\beta})$, $0 < \alpha < \beta < 1$, on $A \in \{A_0, A_0^+, A_0^-\}$ with positively homogeneous score differences.*

Proof. By using Theorem 3.7, any strictly $\mathcal{F}$-consistent scoring function for the functional

$$T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$$

must be of the form (3.3), where in particular ϕ is strictly convex, twice differentiable and ϕ' is bounded. Assume that S has positively homogeneous score differences of some degree $b \in \mathbb{R}$. That means that the function $\Psi \colon (0, \infty) \times \mathsf{A} \times \mathsf{A} \times \mathbb{R} \to \mathbb{R}$ defined by

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

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

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

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 $\phi''(-x_3) = \phi''(-1)x_3^{b-2}$ for any $x_3 > 0$. Due to the strict convexity of ϕ, we need that $\phi''(-1) > 0$. However, for $b \geq 1$ we have $\inf_{x_3 > 0} \phi'(-x_3) = -\infty$, and for $b \leq 1$ we have $\sup_{x_3 > 0} \phi'(-x_3) = \infty$. Hence, ϕ' cannot be bounded. $\qquad\square$

Remark 4.3. The negative result of Proposition 4.2 should be compared with the results of Nolde and Ziegel [43, Theorem C.3] characterizing 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^-. When interpreting $\mathrm{RVaR}_{\alpha,\beta}$ as a risk measure, negative values of RVaR are the more interesting and relevant ones, using our sign convention. Inspecting the proofs of Proposition 4.2 and of Proposition 3.5 (i) one makes the following observation: for $b \geq 1$, Nolde and Ziegel [43] 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 3.5, 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 [21, Theorem 5.2]. 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 [43], 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 consistent scoring functions, one has to be aware that in the presence of non-nested information sets, model mis-specification and/or finite samples, the ranking may depend on the chosen consistent scoring function [46]. In the specific case of $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$, the forecast ranking may depend on the specific choice for the functions g_1, g_2, and ϕ appearing in Theorem 3.3. 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, Gneiting, Jordan and Krüger [12]. 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). The applicability is illustrated in Section 6. Recall that $S_\alpha(x, y) = (\mathbb{1}\{y \leq x\} - \alpha)x - \mathbb{1}\{y \leq x\}y$.

Theorem 5.1. *Let* $0 < \alpha < \beta < 1$. *Any scoring function*

$$S \colon [c_{\min}, c_{\max}]^3 \times \mathbb{R} \to \mathbb{R}$$

of the form (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

$$L_v^1(x_1, y) = (\mathbb{1}\{y \le x_1\} - \alpha)(\mathbb{1}\{v \le x_1\} - \mathbb{1}\{v \le y\}),$$

$$L_v^2(x_2, y) = (\mathbb{1}\{y \le x_2\} - \beta)(\mathbb{1}\{v \le x_2\} - \mathbb{1}\{v \le y\}),$$

$$L_v^3(x_1, x_2, x_3, y) = \frac{1}{\beta - \alpha}\left(\mathbb{1}\{v > x_3\}(S_\alpha(x_1, y) + \alpha y) + \mathbb{1}\{v \le x_3\}(S_\beta(x_2, y) + \beta y)\right) + (\mathbb{1}\{v \le x_3\} - \mathbb{1}\{v \le y\})v,$$

and H_1, H_2 are locally finite measures on $[c_{\min}, c_{\max}]$ and H_3 is a finite measure on $[c_{\min}, c_{\max}]$. 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 $h : [c_{\min}, c_{\max}] \to \mathbb{R}$ can always be written as

$$h(x) = \int (\mathbb{1}\{v \le x\} - \mathbb{1}\{v \le z\})\, dH(v) + C, \quad x \in [c_{\min}, c_{\max}], \tag{5.2}$$

for some locally finite measure H and some $z \in [c_{\min}, c_{\max}]$, $C \in \mathbb{R}$. The function h is strictly increasing if and only if H is strictly positive, i.e., it puts positive mass on all open non-empty 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 $A \subseteq [c_{\min}, c_{\max}]$, where $\mathcal{L}$ is the Lebesgue measure on $\mathbb{R}$.

Using the arguments from Proposition 3.5, it is no loss of generality to show the assertion for a score S such that $\lambda(\beta - \alpha) = -\inf_x \phi'(x) = \sup_x \phi'(x)$ and the one-sided derivatives of g_1, g_2 are bounded from below by $\lambda > 0$.

Then there is a measure H_3 on $[c_{\min}, c_{\max}]$ such that $H_3([c_{\min}, c_{\max}]) = 2\lambda(\beta - \alpha)$, which is strictly positive if and only if ϕ is strictly convex, such that for all $x_3 \in [c_{\min}, c_{\max}]$ we have

$$\phi'(x_3) = \int \mathbb{1}\{v \le x_3\}\, dH_3(v) - \lambda(\beta - \alpha) = \int \left(\mathbb{1}\{v \le x_3\} - \frac{1}{2}\right) dH_3(v).$$

Using Fubini's theorem, we find that

$$\phi(x_3) - \phi(y) = \int (\mathbb{1}\{w \le x_3\} - \mathbb{1}\{w \le y\})\phi'(w)\, dw$$

$$= \int (\mathbb{1}\{w \le x_3\} - \mathbb{1}\{w \le y\}) \int \left(\mathbb{1}\{v \le w\} - \frac{1}{2}\right) dH_3(v)\, dw$$

$$= \int \int (\mathbb{1}\{w \le x_3\} - \mathbb{1}\{w \le y\})\mathbb{1}\{v \le w\}\, dw\, dH_3(v) - \int \frac{1}{2}(x_3 - y)\, dH_3(v)$$

$$= \int \mathbb{1}\{v \le x_3\}(x_3 - v) - \mathbb{1}\{v \le y\}(y - v) - \frac{1}{2}(x_3 - y)\, dH_3(v).$$

By using (3.3), (5.2) and Proposition 3.5, it is straightforward to check that a scoring function of the form (3.3) can be written as in (5.1) with L_v^3 replaced by

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

and locally finite measures $\tilde{H}_1, \tilde{H}_2$ on $[c_{\min}, c_{\max}]$ instead of H_1, H_2 such that $\tilde{H}_i(A) \ge \lambda \mathcal{L}(A)$ for $i = 1, 2$, and for all Borel sets $A \subseteq \mathbb{R}$ and the measure H_3. 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 $H_3([c_{\min}, c_{\max}]) = 2\lambda(\beta - \alpha)$ yields the claim with

$$L_v^3(x_1, x_2, x_3, y) = \frac{1}{2(\beta - \alpha)}(S_\beta(x_2, y) + \beta y + S_\alpha(x_1, y) + \alpha y)$$

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

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 (3.3) with

$$g_1(x) = g_2(x) = \frac{x}{2\beta - 2\alpha} \quad \text{and} \quad \phi(x) = \frac{|x - v|}{2},$$

which renders it a consistent scoring function for $(\text{VaR}_\alpha, \text{VaR}_\beta, \text{RVaR}_{\alpha,\beta})$. The converse statement follows by direct computations. $\qquad\square$
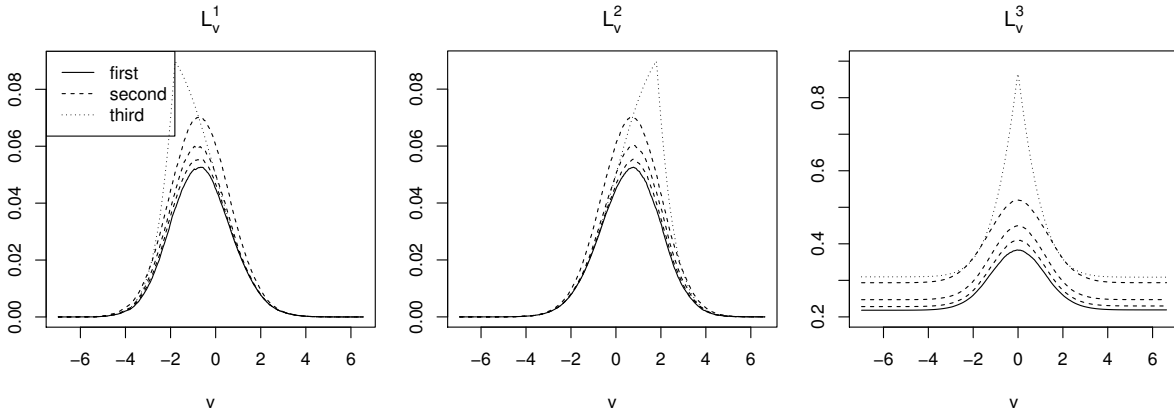
Figure 1: Murphy diagrams for $\alpha = 1 - \beta = 0.1$. Plots of expected elementary scores L_v^1, L_v^2, L_v^3 in terms of v for the three forecasters described in the text. For the second forecaster, the curves correspond to $\sigma = 0.3, 0.5, 0.8$ from bottom to top.
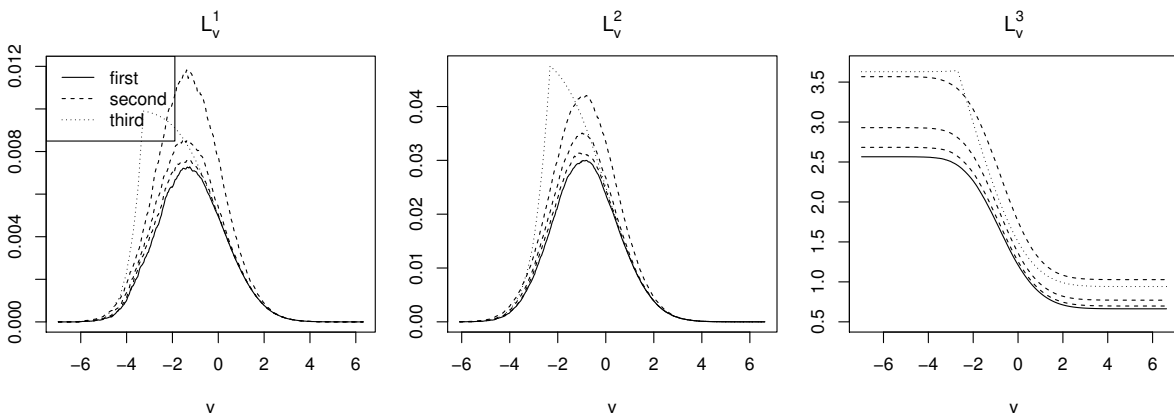

Figure 2: Murphy diagrams for $\alpha = 0.01, \beta = 0.05$. Plots of expected elementary scores L_v^1, L_v^2, L_v^3 in terms of v for the three forecasters described in the text. For the second forecaster, the curves correspond to $\sigma = 0.3, 0.5, 0.8$ from bottom to top.

6 Simulations

This simulation study illustrates the usage of consistent scoring functions for the triplet

$$(\text{VaR}_\alpha, \text{VaR}_\beta, \text{RVaR}_{\alpha,\beta})$$

when comparing the predictive performances of different forecasts for this triplet, e.g., in the context of comparative backtests [43]. We use the scoring functions given in Table 1 and discussed in Example 3.6. The only modification is that in the cases of S_1, S_2, S_3 we additionally scale the functions ϕ' (and therefore also ϕ), working with $\tilde{\phi}'(x_3) = \phi'((\beta - \alpha)x_3)$. This choice has performed better in our simulations. We illustrate the discrimination ability of the suggested scoring functions with a slightly extended version of a simulation example of [28] which has also been considered in [24]. It features a cross-sectional setup. Similar simulation studies can also be performed in an autoregressive time series framework.

Let us first introduce the data generating process. To this end, let $(W_t, Z_t, u_t)'$ be an i.i.d. sequence with a centered Gaussian distribution and diagonal covariance matrix with diagonal entries $(1, \sigma^2, 1)'$. The variables W_t and Z_t will play the role of explanatory variables, and u_t is an unobservable error term. Let $Y_t = W_t + u_t$ be our sequence of observations. Therefore, $Y_t \sim \mathcal{N}(0, 2)$ and $Y_t | W_t \sim \mathcal{N}(W_t, 1)$, while Z_t is completely uninformative since it is independent of Y_t. Suppose we have three different forecasters who provide point forecasts, aiming at correctly specifying $T = (\text{VaR}_\alpha, \text{VaR}_\beta, \text{RVaR}_{\alpha,\beta})$ of the (conditional) distribution of Y_t. The first forecaster has access to the explanatory variables W_t, Z_t and issues correctly specified condi-

tional risk measure forecasts

$$\widehat{T}_t^{(1)} = T(F_{Y_t|W_t,Z_t}) = T(F_{Y_t|W_t})$$
$$= \left(W_t + \Phi^{-1}(\alpha),\ W_t + \Phi^{-1}(\beta),\ W_t - \frac{1}{\beta - \alpha}\big(\varphi(\Phi^{-1}(\beta)) - \varphi(\Phi^{-1}(\alpha))\big) \right)$$

for the time point t, where φ and Φ denote the density and quantile function of the standard normal distribution, respectively. The second forecaster also has access to W_t and Z_t. However, they use a wrong model, issuing (correct) forecasts for $\widetilde{Y}_t = W_t + Z_t + u_t$ rather than for Y_t. That means

$$\widehat{T}_t^{(2)} = T(F_{\widetilde{Y}_t|W_t,Z_t}) = (\widehat{T}_{1,t}^{(1)} + Z_t,\ \widehat{T}_{2,t}^{(1)} + Z_t,\ \widehat{T}_{3,t}^{(1)} + Z_t).$$

The third forecaster is uninformed and makes correct unconditional predictions:

$$\widehat{T}_t^{(3)} = T(F_{Y_t}) = \left(\sqrt{2}\Phi^{-1}(\alpha),\ \sqrt{2}\Phi^{-1}(\beta),\ -\frac{\sqrt{2}}{\beta - \alpha}\big(\varphi(\Phi^{-1}(\beta)) - \varphi(\Phi^{-1}(\alpha))\big) \right).$$

Applying the very definition of (strict) consistency, it holds for any (strictly) consistent scoring function for T that

$$\mathbb{E}(S(\widehat{T}_t^{(1)}, Y_t) \mid W_t, Z_t) \le \mathbb{E}(S(\widehat{T}_t^{(2)}, Y_t) \mid W_t, Z_t)$$

almost surely with a strict inequality in case of strict consistency. Therefore, also

$$\mathbb{E}(S(\widehat{T}_t^{(1)}, Y_t)) \le \mathbb{E}(S(\widehat{T}_t^{(2)}, Y_t)).$$

That means, forecaster 1 should be (strictly) preferred to forecaster 2 under any (strictly) consistent scoring function. Similarly, forecaster 1 should (strictly) outperform forecaster 3 with respect to any (strictly) consistent scoring functions, due to increasing information sets; invoking [31]. Indeed, we have

$$\mathbb{E}(S(\widehat{T}_t^{(1)}, Y_t) \mid W_t) \le \mathbb{E}(S(\widehat{T}_t^{(3)}, Y_t) \mid W_t)$$

almost surely such that $\mathbb{E}(S(\widehat{T}_t^{(1)}, Y_t)) \le \mathbb{E}(S(\widehat{T}_t^{(3)}, Y_t))$ follows with strict inequalities when S is strictly consistent. When comparing forecasters 2 and 3, it is not a priory clear which forecaster is preferred. It will generally depend on the choice of the (strictly) consistent scoring function and on the size of the variance σ^2. Recalling that the limiting case of $\sigma^2 \to 0$ yields forecaster 1, forecaster 2 should be preferred for small σ^2. Their performance should deteriorate as σ^2 increases.

Figures 1 and 2 provide Murphy diagrams of all forecasters computed from a sample of size $N = 100\,000$, providing a good approximation of the population level. They are in line with our theoretical considerations above concerning the ranking of the three forecasts.

We also consider a setup which is closer to the stylistic situation in comparative backtests in a risk management context. To this end, we compare the predictive performances using Diebold–Mariano tests [10] based on the scoring functions in Table 1 (scaled as explained previously). 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 empirical ratio of rejections of the null hypothesis that forecaster i outperforms forecaster j, $i, j \in \{1, 2, 3\}$, $i \ne j$, evaluated in terms of the score S at significance level 0.05. That is, we consider the null hypothesis $\mathbb{E}(S(\widehat{T}_t^{(i)}, Y_t)) \le \mathbb{E}(S(\widehat{T}_t^{(j)}, Y_t))$ for all $t = 1, \ldots, N$, or in short, $i \preceq j$. 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 scoring functions used. For a comparison of forecaster 1 and forecaster 2, the situation is more interesting: 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) is very similar for S_1, S_2 and S_3 (around 0.864 to 0.873). On the other hand, S_4 achieves a considerably higher power of 0.956. The situation described in the right panel of Table 2 considering the

H_0	S_1	S_2	S_3	S_4	H_0	S_1	S_2	S_3	S_4
$1 \preceq 2$	0	0	0	0	$1 \preceq 2$	0	0	0	0
$2 \preceq 1$	0.864	0.864	0.873	0.956	$2 \preceq 1$	0.675	0.671	0.670	0.522
$1 \preceq 3$	0	0	0	0	$1 \preceq 3$	0	0	0	0
$3 \preceq 1$	1.000	1.000	1.000	1.000	$3 \preceq 1$	0.992	0.992	0.994	0.817
$2 \preceq 3$	0	0	0	0	$2 \preceq 3$	0	0	0	0.002
$3 \preceq 2$	0.999	0.999	0.990	0.996	$3 \preceq 2$	0.740	0.742	0.768	0.258

Table 2: Power of Diebold–Mariano tests at significance level 0.05 for the scoring functions in Table 1 (suitably scaled) 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 $i \preceq j$ means that $\mathbb{E}(S(\widehat{T}_t^{(i)}, Y_t)) \leq \mathbb{E}(S(\widehat{T}_t^{(j)}, Y_t))$ for all $t = 1, \dots, N$ for the scoring function specified in the column label. We chose $\sigma^2 = 0.5^2$ for the forecaster 2.

parameter choice $\alpha = 0.01$ and $\beta = 0.05$ leads to a different situation. The most obvious observation is that the power is lower than in the symmetric situation depicted in the left panel for all null hypotheses, respectively. This intuitively makes sense since differences in the tail behavior are more challenging to detect in comparison to differences in the behavior of the central region of the distribution. Second, we can see that the power of the scores S_1, S_2 and S_3 is again very similar for all situations, whereas the score S_4 performs apparently worse. This can be seen most strikingly for the null $3 \preceq 2$: the power of the scores S_1, S_2 and S_3 is between 0.740 and 0.768, whereas S_4 yields a power of only 0.258. However, as mentioned above, for a comparison between forecaster 2 and 3, it is also not possible to establish a general ranking for all consistent scoring functions. In line with this, the dependence of the ranking on the choice of the score is reflected in the difference in power. A more detailed study and comparison of other scoring functions and other situations is deferred to future work.

7 Implications for regression

After illustrating the usage of consistent scoring functions in forecast comparison and comparative back-testing in Section 6, 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 generalized 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 $(W_t, Y_t)_{t\in\mathbb{N}}$ be a time series with the usual notation that Y_t denotes some real-valued response variable and W_t is a d-dimensional vector of regressors. Let $\Theta \subseteq \mathbb{R}^k$ be some parameter space and let $M \colon \mathbb{R}^d \times \Theta \to \mathbb{R}^3$ be a parametric model for $T = (\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$, $0 < \alpha < \beta < 1$. We assume a correct model specification, that is, we assume that there is a unique $\theta_0 \in \Theta$ such that

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

where $F_{Y_t|W_t}$ denotes the conditional distribution of Y_t given W_t. That means, $M(W_t, \theta_0)$ 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 suppose the sequence $(W_t, Y_t)_{t\in\mathbb{N}}$ satisfies certain mixing conditions [54, Corollary 3.48] (of which a special case is independence). Then one obtains under additional moment conditions that, as $n \to \infty$,

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

It is essentially this law of large numbers result which allows for consistent parameter estimation with the empirical M-estimator $\widehat{\theta}_n = \arg\min_{\theta \in \Theta} n^{-1} \sum_{t=1}^{n} S(M(W_t, \theta), Y_t)$; see, e.g., [11, 33, 43, 50] for details.

In summary, we can see that the complication of this procedure is that one needs to model the components VaR_α and 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 serially dependent, though mixing, and non-stationary data. One only needs the *semiparametric stationarity* specified through equation (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, $\alpha \in (0, \frac{1}{2})$, 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 Bassett [35] – cf. [49] – 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. By using the notation from Section 7.1, an M-estimator for $\mathrm{RVaR}_{\alpha,\beta}$ is given by

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

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 W_t and only modelling the intercept part), it is not clear how to use it in a more interesting regression context, where one is actually interested in the *conditional* distribution of Y_t given W_t rather than the unconditional distribution of Y_t. Moreover, since this approach uses the order-statistics of the entire sample $Y_1, \ldots, Y_n$ to implicitly estimate the α- and β-quantile, it requires that these quantiles be constant in time. Hence, heteroscedasticity (in time) can lead to problems, even if $\mathrm{RVaR}_{\alpha,\beta}$ is constant in time.

A second approach is described, for example, in [47, 48] 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

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

For each $\theta \in \Theta$, define the residuals $\varepsilon_t(\theta) := Y_t - m(W_t, \theta)$ and the absolute residuals $r_t(\theta) := |\varepsilon_t(\theta)|$. 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 M-estimator is defined via

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

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 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 M-estimator via

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

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, as outlined above, this procedure can lead to problems in the presence of heteroscedasticity or general non-stationarity of the error distribution if the conditional VaR_α and VaR_β of Y_t given W_t depend on W_t. 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 Section 7.1 does not rely on the usage of order-statistics and it can in general deal with heteroscedasticity. The only degree of "stationarity" is required through (7.1). Especially, stationarity is deemed too strong an assumption in the context of financial data; see [9].

Finally, we would like to remark that there are further procedures belonging to the field of TLS. For instance, Atkinson and Cheng [4] propose an adaptive procedure where the trimming parameter is data driven; see also [7]. 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 [32] introduced the famous *Huber loss* $S(x, y) = \rho(x - y)$, where $\rho(t) = \frac{1}{2}t^2$ for $|t| \leq k$ and $\rho(t) = k|t| - \frac{1}{2}k^2$ for $|t| > k$. Huber argues that "the corresponding [M-]estimator is related to Winsorizing" [32, p. 79]. What obtained significantly less attention – maybe due to its 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 $|t| \leq k$ 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

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

Assuming that F is continuous with density f for the sake of the simplicity of the argument, the corresponding first-order condition for a minimum of the expected score $\bar{S}_{k_1,k_2}(x, F)$ is equivalent to

$$x = \frac{1}{F(k_2 - x) - F(k_1 - x)} \int_{k_1 - x}^{k_2 - x} y f(y) \, \mathrm{d}y.$$

Now, a suggestion similar to [47, 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 in this paper at (3.3).

A Appendix

We present a list of assumptions used in Section 3. For more details about their interpretations and implications, please see [21] where they were originally introduced.

Assumption (V1). $\mathcal{F}$ is convex and for every $x \in \text{int}(A)$ there are $F_1, \ldots, F_{k+1} \in \mathcal{F}$ such that

$$0 \in \text{int}(\text{conv}(\{\bar{V}(x, F_1), \ldots, \bar{V}(x, F_{k+1})\})).$$

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

Assumption (V3). The map $\bar{V}(\cdot, F)$ is continuously differentiable for every $F \in \mathcal{F}$.

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

$$\partial_l \bar{V}_l(t, F_1) = \partial_l \bar{V}_l(t, F_2) \quad \text{for all } l \in \{1, \ldots, k\} \setminus \{r\},$$
$$\partial_r \bar{V}_r(t, F_1) \neq \partial_r \bar{V}_r(t, F_2).$$

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

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

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 $t = T(F) \in \text{int}(A)$.

References

[1] C. Acerbi and B. Székely, Backtesting expected shortfall, *Risk Mag.* (2014), 1–33.

[2] C. Acerbi and B. Székely, General properties of backtestable statistics, preprint (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2905109.

[3] P. Artzner, F. Delbaen, J.-M. Eber and D. Heath, Coherent measures of risk, *Math. Finance* **9** (1999), no. 3, 203–228.

[4] A. C. Atkinson and T.-C. Cheng, Computing least trimmed squares regression with the forward search, *Statist. Comput.* **9** (1999), no. 4, 251–263.

[5] S. Barendse, Efficiently weighted estimation of tail and interquartile expectations, preprint (2020), https://dx.doi.org/10.2139/ssrn.2937665.

[6] J. R. Brehmer, *Elicitability and its application in risk management*, Master's thesis, University of Mannheim, 2017.

[7] 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** (2018), no. 4, 559–587.

[8] R. Cont, R. Deguest and G. Scandolo, Robustness and sensitivity analysis of risk measurement procedures, *Quant. Finance* **10** (2010), no. 6, 593–606.

[9] M. H. A. Davis, Verification of internal risk measure estimates, *Stat. Risk Model.* **33** (2016), no. 3–4, 67–93.

[10] F. X. Diebold and R. S. Mariano, Comparing predictive accuracy, *J. Bus. Econom. Statist.* **13** (1995), 253–263.

[11] T. Dimitriadis, T. Fissler and J. F. Ziegel, The efficiency gap, preprint (2020), https://arxiv.org/abs/2010.14146.

[12] 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** (2016), no. 3, 505–562.

[13] P. Embrechts, H. Liu, T. Mao and R. Wang, Quantile-based risk sharing with heterogeneous beliefs, *Math. Program.* **181** (2020), no. 2, 319–347.

[14] P. Embrechts, H. Liu and R. Wang, Quantile-based risk sharing, *Oper. Res.* **66** (2018), no. 4, 936–949.

[15] P. Embrechts, G. Puccetti, L. Rüschendorf, R. Wang and A. Beleraj, An academic response to Basel 3.5, *Risks* **2** (2014), 25–48.

[16] P. Embrechts, B. Wang and R. Wang, Aggregation-robustness and model uncertainty of regulatory risk measures, *Finance Stoch.* **19** (2015), no. 4, 763–790.

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

[18] J. Engelberg, C. F. Manski and J. Williams, Comparing the point predictions and subjective probability distributions of professional forecasters, *J. Bus. Econom. Statist.* **27** (2009), no. 1, 30–41.

[19] T. Fissler, *On higher order elicitability and some limit theorems on the poisson and Wiener space*, PhD thesis, University of Bern, 2017.

[20] T. Fissler, R. Frongillo, J. Hlavinová and B. Rudloff, Forecast evaluation of quantiles, prediction intervals, and other set-valued functionals, *Electron. J. Stat.* **15** (2021), no. 1, 1034–1084.

[21] T. Fissler and J. F. Ziegel, Higher order elicitability and Osband's principle, *Ann. Statist.* **44** (2016), no. 4, 1680–1707.

[22] T. Fissler and J. F. Ziegel, Order-sensitivity and equivariance of scoring functions, *Electron. J. Stat.* **13** (2019), no. 1, 1166–1211.

[23] T. Fissler and J. F. Ziegel, Correction note: Higher order elicitability and Osband's principle, *Ann. Statist.* **49** (2021), no. 1, 614.

[24] T. Fissler, J. F. Ziegel and T. Gneiting, Expected shortfall is jointly elicitable with value-at-risk: Implications for backtesting, *Risk Mag.* (2016), 58–61.

[25] R. Frongillo and I. Kash, Elicitation complexity of statistical properties, *Biometrika* (2020), DOI 10.1093/biomet/asaa093.

[26] R. Giacomini and H. White, Tests of conditional predictive ability, *Econometrica* **74** (2006), no. 6, 1545–1578.

[27] T. Gneiting, Making and evaluating point forecasts, *J. Amer. Statist. Assoc.* **106** (2011), no. 494, 746–762.

[28] T. Gneiting, F. Balabdaoui and A. E. Raftery, Probabilistic forecasts, calibration and sharpness, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** (2007), no. 2, 243–268.

[29] T. Gneiting and A. E. Raftery, Strictly proper scoring rules, prediction, and estimation, *J. Amer. Statist. Assoc.* **102** (2007), no. 477, 359–378.

[30] F. R. Hampel, A general qualitative definition of robustness, *Ann. Math. Statist.* **42** (1971), 1887–1896.

[31] H. Holzmann and M. Eulert, The role of the information set for forecasting—with applications to risk management, *Ann. Appl. Stat.* **8** (2014), no. 1, 595–621.

[32] P. J. Huber, Robust estimation of a location parameter, *Ann. Math. Statist.* **35** (1964), 73–101.

[33] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, 2nd ed., John Wiley & Sons, Hoboken, 2009.

[34] R. Koenker, *Quantile Regression*, Cambridge University, Cambridge, 2005.

[35] R. Koenker and G. Bassett, Jr., Regression quantiles, *Econometrica* **46** (1978), no. 1, 33–50.

[36] S. Kou, X. Peng and C. C. Heyde, External risk measures and Basel accords, *Math. Oper. Res.* **38** (2013), no. 3, 393–417.

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

[38] V. Krätschmer, A. Schied and H. Zähle, Comparative and qualitative robustness for law-invariant risk measures, *Finance Stoch.* **18** (2014), no. 2, 271–295.

[39] N. Lambert, D. M. Pennock and Y. Shoham, Eliciting properties of probability distributions, in: *Proceedings of the 9th ACM Conference on Electronic Commerce*, ACM, New York (2008), 129–138.

[40] G. Lugosi and S. Mendelson, Robust multivariate mean estimation: the optimality of trimmed mean, *Ann. Statist.* **49** (2021), no. 1, 393–410.

[41] A. H. Murphy and H. Daan, Forecast evaluation, in: *Probability, Statistics and Decision Making in the Atmospheric Sciences*, Westview Press, Boulder (1985), 379–437.

[42] W. K. Newey and J. L. Powell, Asymmetric least squares estimation and testing, *Econometrica* **55** (1987), no. 4, 819–847.

[43] N. Nolde and J. F. Ziegel, Elicitability and backtesting: Perspectives for banking regulation, *Ann. Appl. Stat.* **11** (2017), no. 4, 1833–1874.

[44] K. H. Osband, *Providing incentives for better cost forecasting*, PhD thesis, University of California, Berkeley, 1985.

[45] A. J. Patton, Data-based ranking of realised volatility estimators, *J. Econometrics* **161** (2011), no. 2, 284–303.

[46] A. J. Patton, Comparing possibly misspecified forecasts, *J. Bus. Econom. Statist.* **38** (2020), no. 4, 796–809.

[47] P. Rousseeuw, Least median of squares regression, *J. Amer. Statist. Assoc.* **79** (1984), no. 388, 871–880.

[48] P. Rousseeuw, Multivariate estimation with high breakdown point, in: *Mathematical Statistics and Applications*, Reidel, Dordrecht (1985), 283–297.

[49] D. Ruppert and R. J. Carroll, Trimmed least squares estimation in the linear model, *J. Amer. Statist. Assoc.* **75** (1980), no. 372, 828–838.

[50] A. W. van der Vaart, *Asymptotic Statistics*, Camb. Ser. Stat. Probab. Math. 3, Cambridge University, Cambridge, 1998.

[51] R. Wang and Y. Wei, Risk functionals with convex level sets, *Math. Finance* **30** (2020), no. 4, 1337–1367.

[52] S. Wang, Insurance pricing and increased limits ratemaking by proportional hazards transforms, *Insurance Math. Econom.* **17** (1995), no. 1, 43–54.

[53] S. Weber, Distribution-invariant risk measures, information, and dynamic consistency, *Math. Finance* **16** (2006), no. 2, 419–441.

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

[55] M. E. Yaari, The dual theory of choice under risk, *Econometrica* **55** (1987), no. 1, 95–115.

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

[57] J. F. Ziegel, Coherence and elicitability, *Math. Finance* **26** (2016), no. 4, 901–918.

[58] Bank for International Settlements, Consultative Document: Fundamental review of the trading book: Outstanding issues, 2014.