



## Decision Support

## Robust elicitable functionals

Kathleen E. Miao<sup>ID\*</sup>, Silvana M. Pesenti<sup>ID</sup>

Department of Statistical Sciences, University of Toronto, Canada



## ARTICLE INFO

## Keywords:

Elicitability

Kullback–Leibler divergence

Model uncertainty

Risk measures

Distributional robustness

## ABSTRACT

Elicitable functionals and (strictly) consistent scoring functions are of interest due to their utility of determining (uniquely) optimal forecasts, and thus the ability to effectively backtest predictions. However, in practice, assuming that a distribution is correctly specified is too strong a belief to reliably hold. To remediate this, we incorporate a notion of statistical robustness into the framework of elicitable functionals, meaning that our robust functional accounts for “small” misspecifications of a baseline distribution. Specifically, we propose a robustified version of elicitable functionals by using the Kullback–Leibler divergence to quantify potential misspecifications from a baseline distribution. We show that the robust elicitable functionals admit unique solutions lying at the boundary of the uncertainty region, and provide conditions for existence and uniqueness. Since every elicitable functional possesses infinitely many scoring functions, we propose the class of b-homogeneous strictly consistent scoring functions, for which the robust functionals maintain desirable statistical properties. We show the applicability of the robust elicitable functional in several examples: in a reinsurance setting and in robust regression problems.

## 1. Introduction

Risk measures are tools used to quantify profits and losses of financial assets. A risk measure maps losses, stemming from, e.g. historical data or a distributional output, to a real number characterising its riskiness or in an insurance setting, e.g. the amount of assets required to be retained as reserves. However, in many situations, data or distributional information of modelled losses may be flawed for several reasons — they may be out-of-date, sparse, or unreliable due to errors. As such, it is of interest to relax the assumption that the underlying distribution is correctly specified. In the literature, this is referred to as distributional robustness, and is often approached via a worst-case risk measure; indicatively see [Bernard et al. \(2024\)](#), [Cai et al. \(2024\)](#) and [Ghaoui et al. \(2003\)](#). For a set of probability measures  $\mathcal{Q}$  — the so-called uncertainty set — a worst-case risk measure evaluated at a random variable  $Y$  is typically defined by

$$\sup_{\mathbb{Q} \in \mathcal{Q}} R(F_Y^{\mathbb{Q}}), \quad (1)$$

where  $R$  is a law-invariant risk measure and  $F_Y^{\mathbb{Q}}$  is the cumulative distribution function (cdf) of  $Y$  under  $\mathbb{Q}$ . Thus, a worst-case risk measure is the largest value the risk measure can attain over a predetermined set of cdfs.

From a statistical perspective, the notion of elicitability of a law invariant functional is of interest as it yields a natural backtesting procedure ([Nolde & Ziegel, 2017](#)). Of course, risk measures are functionals,

and specific risk measures, such as the Value-at-Risk (VaR), are elicitable. An elicitable risk measure can be represented as the minimiser of the expectation of a suitable scoring function ([Gneiting, 2011](#)). Thus, for an elicitable risk measure  $R$  with corresponding scoring function  $S : \mathcal{A} \times \mathbb{R} \rightarrow [0, \infty)$ ,  $\mathcal{A} \subseteq \mathbb{R}$ , the worst-case risk measure given in (1) can equivalently be written as

$$\sup_{\mathbb{Q} \in \mathcal{Q}} \operatorname{argmin}_{z \in \mathcal{A}} \int S(z, y) dF_Y^{\mathbb{Q}}(y). \quad (2)$$

The worst-case risk measure framework yields a worst-case cdf, that is, the one attaining the supremum in (1) and (2). This worst-case cdf may be interpreted as the distribution of the losses when the most adverse (with respect to the risk measure) probability measure in  $\mathcal{Q}$  materialises. Worst-case values, also called risk bounds, have been studied extensively for various choices of uncertainty set and risk measure, see [Rüschedendorf et al. \(2024\)](#) for an extensive overview.

Alternative to the worst-case methodology, we propose leveraging distributional robustness and elicitability to create a robust elicitable functional (REF). Motivated by this, we interchange the supremum and argmin in (2) and define the REF as

$$\bar{R}^S(Y) := \operatorname{argmin}_{z \in \mathcal{A}} \sup_{\mathbb{Q} \in \mathcal{Q}_e} \int S(z, y) dF_Y^{\mathbb{Q}}(y), \quad (3)$$

where in the exposition, we consider an uncertainty set characterised by the Kullback–Leibler (KL) divergence. In contrast to worst-case risk

\* Corresponding author.

E-mail addresses: [k.miao@mail.utoronto.ca](mailto:k.miao@mail.utoronto.ca) (K.E. Miao), [silvana.pesenti@utoronto.ca](mailto:silvana.pesenti@utoronto.ca) (S.M. Pesenti).

measures, the REF depends not only on the choice of risk measure but more specifically on the scoring function that elicits the risk measure, see also Section 3 for a detailed discussion. The optimal measure attaining the inner problem in (3) can be interpreted as the, in KL divergence, largest admissible expected score; the extremal score. As the REF is the minimiser of the extremal score, it is indeed a “robust” risk functional. Contrary to the worst-case risk-measure, which is always larger than the baseline risk measure, the REF, which is the minimiser of the extremal score, can be smaller or larger than the baseline risk measure. From a computational perspective, if the inner optimisation problem of (3) over the space of probability measures can be solved semi-analytically, the outer problem over reals can typically be tackled using classical optimisation techniques.

The benefits of elicitable functionals has been well justified in the domains of statistics and risk management. Elicitable functionals and (strictly) consistent scoring functions are of interest due to their utility of determining (uniquely) optimal forecasts, and thus the ability to effectively backtesting predictions, see e.g. Fissler and Ziegel (2016) and Gneiting (2011). The literature on elicability in risk management is extensive, with many authors arguing for its importance (He et al., 2022; Ziegel, 2016). The characterisation of elicitable convex and coherent risk measures for example has been studied in Bellini and Bignozzi (2015), while Fissler et al. (2016) showed that Expected Shortfall (ES) is jointly elicitable with VaR. Embrechts et al. (2021) explored the connection between elicability and Bayes risks, and showed that entropic risk measures are the only risk measures that have both properties. In the related (Fissler & Pesenti, 2023), authors study sensitivity measures tailored for elicitable risk measures. Robustness to distributional uncertainty is of concern in risk management and has been explored in the literature, see e.g. Embrechts et al. (2013) for approximation of worst-case risk measures using the Rearrangement Algorithm, Pesenti et al. (2016) for a discussion on distributional robustness of distortion risk measures, and Embrechts et al. (2015) for the aggregation robustness of VaR and ES.

Aside from its attention in DRO and the model assessment literature, the KL divergence implies that all alternative distributions have the same support as the reference distribution. This is of particular relevance in applications such as insurance settings where the loss is bounded from above and below, or in regression settings that include categorical, likert or proportional data. We refer to Sections 4.2 and 5 for further details. Distributionally robust optimisation (DRO) studies decision models where the decision maker has only partial distributional knowledge of the model at hand. These constrained optimisation problems take the form of a minimax problem, akin to Eq. (3). A method of characterising uncertainty is to use the KL divergence, as introduced in Kullback and Leibler (1951). In the realm of DRO and minimax problems, KL uncertainty is often considered, e.g. Hu and Hong (2013) study DRO problems with ambiguity sets constrained by the KL divergence, and Calafiori (2007) finds mean-risk optimal robust portfolios under distributional uncertainty characterised by the KL divergence. The minimax problem we propose in Eq. (3) fits into the general framework of the DRO literature. With the REF methodology, we not only solve the inner problem semi-analytically but we also show existence and uniqueness of the outer problem.

We extend the REF to  $k$ -elicitable functionals, and thus obtain jointly robust VaR and ES pairs. In this setting, we obtain a probability measure that attains extremal value simultaneously for VaR and ES. This is in contrast to the classical framework of worst-case risk measures, which yields different worst-case cdfs for VaR and ES. As the proposed REF depends on the choice of scoring function, we discuss the families of homogeneous scoring function that naturally lead to visualisations of the REFs via Murphy diagrams. We showcase the applicability of the REF on a simulated reinsurance example and connect the framework to robust regression, where we explore a data driven example. The key novelty and findings of this work is the distributionally robust elicitable functional, the REF, given in

(3), which is an alternative to the classical worst-case risk measure by taking the minimiser of a worst-case expected score. We solve the inner optimisation problem of (3) and obtain a semi-closed solution of the extremal probability measure. Furthermore, we prove existence and uniqueness of the REF, that is, the solution to the double optimisation problem (3).

This paper is organised as follows; in Section 2, we define the robust elicitable functional and state our main results — solving the inner problem of (3) and establishing existence and uniqueness of the REF. In Section 3, we address the question of choosing a scoring function by considering families of parameterised scoring functions, showing that these families retain useful properties of the functional, and illustrate the behaviour of the REF in Murphy diagrams. We extend our results to higher order elicitable risk measures such as the (VaR, ES) pair in Section 4, and consider an application of the robust (VaR, ES) in reinsurance. Finally, we explore the connection of the REF to robust regression in Section 5.

## 2. Elicability and robust elicitable functionals

This section provides the necessary notation and recalls the notion of elicability. We further define the robust elicitable functional and prove its existence and uniqueness.

### 2.1. Robust elicitable functionals

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, where we interpret  $\mathbb{P}$  as the baseline probability measure. The baseline probability could, for example, be estimated from data or informed by expert knowledge. When not specified, expectations and distribution are taken with respect to the baseline probability  $\mathbb{P}$ . Let  $\mathcal{L}^\infty := \mathcal{L}^\infty(\Omega, \mathcal{F}, \mathbb{P})$  be the space of essentially bounded random variables (rvs) on  $(\Omega, \mathcal{F}, \mathbb{P})$  and denote by  $\mathcal{M}^\infty$  the corresponding class of cumulative distribution functions (cdfs), i.e.  $\mathcal{M}^\infty := \{F \mid F(y) = \mathbb{P}(Y \leq y), Y \in \mathcal{L}^\infty\}$ .

We denote by  $R : \mathcal{M}^\infty \rightarrow \mathcal{A}$ ,  $\mathcal{A} \subseteq \mathbb{R}$ , a potentially set-valued, law-invariant, and non-constant functional, which is a mapping from the set of admissible cdfs,  $\mathcal{M}^\infty$ , to the set of allowed predictions  $\mathcal{A}$ , also called action domain. We assume that when evaluating the functional  $R(Y)$  that  $Y$  is non-degenerate. As  $R$  is law-invariant, it can equivalently be defined as  $R : \mathcal{L}^\infty \rightarrow \mathcal{A}$ , by setting  $R(Y) := R(F_Y)$ , where  $F_Y$  is the cdf of  $Y \in \mathcal{L}^\infty$ . Denote the (left-)quantile of  $Y$  as  $F_Y^{-1}(q) := \inf\{y \in \mathbb{R} : \mathbb{P}(Y \leq y) \geq q\}$ . We further recall the definition of the moment generating function (mgf) of a rv  $Y \in \mathcal{L}^\infty$  evaluated at  $t \in \mathbb{R}$  as  $M_Y(t) := \mathbb{E}[e^{tY}]$ , and its cumulant generating function (cgf)  $K_Y(t) := \log(M_Y(t))$ .

Throughout we work with elicitable functionals, for this, we first recall the definition of scoring functions and elicability. In statistics, scoring functions are used to assess the accuracy of a (statistical) estimate or prediction to the true value. Formally,  $S(z, y)$  is a mapping from a prediction  $z$  and a realisation  $y$  of  $Y$  to the non-negative reals. By convention, we view  $S$  as a penalty for the forecast  $z$  compared to the realised event  $y$ , where smaller values represent more accurate predictions.

**Definition 1 (Scoring Functions and Elicitability).** A scoring function is a measurable function  $S : \mathcal{A} \times \mathbb{R} \rightarrow [0, \infty)$ . A scoring function may satisfy the following properties:

- (i) A scoring function  $S$  is **consistent** for a functional  $R$ , if

$$\int S(t, y) dF(y) \leq \int S(z, y) dF(y), \quad (4)$$

for all  $F \in \mathcal{M}^\infty$ ,  $t \in R(F)$ , and all  $z \in \mathcal{A}$ .

- (ii) A scoring function  $S$  is **strictly consistent** for a functional  $R$ , if it is consistent and Eq. (4) holds with equality only if  $z \in R(F)$ .
- (iii) A functional  $R$  is **elicitable**, if there exists a strictly consistent scoring function  $S$  for  $R$ .

An elicitable functional  $R$  admits the representation

$$R(Y) = \operatorname{argmin}_{z \in A} \int S(z, y) dF_Y(y), \quad (5)$$

for all  $F_Y \in \mathcal{M}^\infty$  and where  $S$  is any strictly consistent scoring function for  $R$ . We note that  $R(Y)$  can be an interval, such as in the case of the quantile, and we denote the image of  $R$  by  $\operatorname{Im} R \subset A$ . In the case that  $R(Y)$  is an interval, we denote the interval as  $[R(Y)^l, R(Y)^u]$ , with lower endpoint  $R(Y)^l$  and upper endpoint  $R(Y)^u$ , respectively. We call  $R(Y)$  unique, if the argmin is a singleton. Many statistical functionals and risk measures are elicitable, including the mean, median, quantiles, expectiles, see e.g. Bellini and Bigazzi (2015), and Gneiting (2011). Thus, an elicitable risk measure is the best prediction  $z$  that minimises the expected score under the baseline probability  $\mathbb{P}$ .

The definition of elicability, in particular Eq. (5), assumes knowledge of the true distribution of  $Y$ . However, under distributional ambiguity — that is, uncertainty on the distribution of  $Y$  — one may wish consider a robust functional. A classical choice to model distributional deviations is the Kullback–Leibler (KL) divergence. The KL divergence has been extensively used in model assessment; indicatively see Blanchet et al. (2019), Glasserman and Xu (2014), Lassance and Vrins (2023) and Pesenti et al. (2019). For a probability measure  $\mathbb{Q}$  on  $(\Omega, \mathfrak{F})$ , the KL divergence from  $\mathbb{Q}$  to  $\mathbb{P}$  is defined as

$$D_{KL}(\mathbb{Q} \parallel \mathbb{P}) := \mathbb{E}^{\mathbb{P}} \left[ \frac{d\mathbb{Q}}{d\mathbb{P}} \log \left( \frac{d\mathbb{Q}}{d\mathbb{P}} \right) \right],$$

if  $\mathbb{Q}$  is absolutely continuous with respect to  $\mathbb{P}$  and  $+\infty$  otherwise. We denote by  $\mathbb{E}^{\mathbb{Q}}[\cdot]$  the expected value under  $\mathbb{Q}$ , and for simplicity write  $\mathbb{E}[\cdot] := \mathbb{E}^{\mathbb{P}}[\cdot]$ , when considering the baseline probability measure  $\mathbb{P}$ . We use the KL divergence to describe an uncertainty set and propose the following robust elicitable functionals.

**Definition 2 (Robust Elicitable Functional).** Let  $R$  be an elicitable functional with strictly consistent scoring function  $S : A \times \mathbb{R} \rightarrow [0, \infty)$  and let  $\varepsilon \geq 0$ . Then we define the **robust elicitable functional** (REF), for  $Y \in \mathcal{L}^\infty$ , by

$$\bar{R}^S(Y) := \operatorname{argmin}_{z \in A} \sup_{\mathbb{Q} \in \mathcal{Q}_\varepsilon} \mathbb{E}^{\mathbb{Q}}[S(z, Y)], \quad (6)$$

where the uncertainty set  $\mathcal{Q}_\varepsilon$  is given as

$$\mathcal{Q}_\varepsilon := \{\mathbb{Q} \mid \mathbb{Q} \ll \mathbb{P}, \text{ and } D_{KL}(\mathbb{Q} \parallel \mathbb{P}) \leq \varepsilon\}. \quad (7)$$

The uncertainty set  $\mathcal{Q}_\varepsilon$  contains all probabilities that have a KL divergence to  $\mathbb{P}$  less than or equal to  $\varepsilon$ . For simplicity, we omit the dependence of  $\mathcal{Q}_\varepsilon$  on  $\mathbb{P}$  as the baseline is fixed throughout. The parameter  $\varepsilon \geq 0$  is a tolerance distance that quantifies the distance to the baseline probability  $\mathbb{P}$ . Clearly, if  $\varepsilon = 0$ , only the baseline measure is considered and we recover the traditional definition of elicitable functionals. The uncertainty set  $\mathcal{Q}_\varepsilon$  is compact under the topology of weak convergence, for all  $\varepsilon \in (0, \infty)$ , see van Erven and Harremoës (2012). From (6), we see that the REF is the best prediction  $z$ , that minimises the expected score if the universe chose the most adverse distribution of  $Y$ . Thus, the  $\bar{R}^S(Y)$  may be larger or smaller than  $R(Y)$ , hence it is not a worst-case functional in the sense of (2), which is always larger than  $R(Y)$ .

When  $Y$  has a positive probability mass at its essential supremum, then for a too large tolerance distance, as specified in the next proposition, the inner problem in (6) becomes degenerate. The next two results make this precise.

**Lemma 1.** Let  $Y$  be a rv satisfying  $\mathbb{P}(Y = \operatorname{ess sup} Y) =: p > 0$  and define the probability measure

$$\frac{d\mathbb{Q}^s}{d\mathbb{P}} := \frac{e^{sY}}{\mathbb{E}[e^{sY}]}.$$

Then the following limits hold

$$\lim_{s \rightarrow 0} D_{KL}(\mathbb{Q}^s \parallel \mathbb{P}) = 0 \quad \text{and} \quad \lim_{s \rightarrow +\infty} D_{KL}(\mathbb{Q}^s \parallel \mathbb{P}) = \log\left(\frac{1}{p}\right).$$

**Proof.** The fact that  $\lim_{s \rightarrow 0} D_{KL}(\mathbb{Q}^s \parallel \mathbb{P}) = 0$  follows as  $\lim_{s \rightarrow 0} \frac{d\mathbb{Q}^s}{d\mathbb{P}} = 1$ . For the second limit, we redefine the rv  $Y$  via

$$Y(\omega) = \begin{cases} W(\omega) & \text{on } \omega \in A^c \\ \bar{y} := \operatorname{ess sup} Y & \text{on } \omega \in A, \end{cases}$$

where  $W := Y \mathbb{1}_{\{Y < \operatorname{ess sup} Y\}}$ ,  $\mathbb{P}(Y \in A) = p$ , and  $A^c$  denotes the complement of  $A$ . By setting  $W = 0$  on  $A$ , we have that  $\mathbb{E}[e^{sW}] =: M_W(s)$  is the mgf of  $W$ . For  $\omega \in A^c$ , define  $v := W(\omega)$  then we obtain

$$\frac{d\mathbb{Q}^s}{d\mathbb{P}}(\omega) = \frac{e^{sv}}{M_W(s) + pe^{s\bar{y}}} = \frac{1}{M_{W-v}(s) + pe^{s(\bar{y}-v)}}.$$

As  $M_{W-v}(s) \geq 0$  and  $\bar{y} > v$ , we have that  $\lim_{s \rightarrow +\infty} \frac{d\mathbb{Q}^s}{d\mathbb{P}} = 0$   $\mathbb{P}$ -a.s. on  $A^c$ . For  $\omega \in A$ , we have

$$\frac{d\mathbb{Q}^s}{d\mathbb{P}}(\omega) = \frac{e^{s\bar{y}}}{M_W(s) + pe^{s\bar{y}}} = \frac{1}{M_{W-\bar{y}}(s) + p}.$$

Since  $\bar{y} > W$   $\mathbb{P}$ -a.s., it holds that

$$\lim_{s \rightarrow +\infty} M_{W-\bar{y}}(s) = \mathbb{E} \left[ \lim_{s \rightarrow +\infty} e^{s(W-\bar{y})} \right] = 0,$$

and  $\lim_{s \rightarrow +\infty} \frac{d\mathbb{Q}^s}{d\mathbb{P}} = \frac{1}{p}$   $\mathbb{P}$ -a.s. on  $A$ . Thus, we conclude that  $\lim_{s \rightarrow +\infty} D_{KL}(\mathbb{Q}^s \parallel \mathbb{P}) = \log\left(\frac{1}{p}\right)$ .  $\square$

This lemma shows that if  $Y$  takes positive probability mass at its essential supremum, the KL-divergence converges to  $\log\left(\frac{1}{p}\right)$ , which provides an upper limit on the choice of tolerance distance  $\varepsilon$ , as detailed in the next assumption. We note that if  $Y$  does not have a positive probability mass at its essential supremum, then there are no restrictions on the tolerance distance.

**Assumption 1 (Maximal Kullback–Leibler Distance).** Define  $p(z) := \mathbb{P}(S(z, Y) = \operatorname{ess sup} S(z, Y))$ . Then one of the following holds:

- (i) If  $p(z) = 0$  for all  $z \in A$ , then the tolerance distance satisfies  $\varepsilon \in [0, \infty)$ .
- (ii) If  $p(z) > 0$  for some  $z \in A$ , then the tolerance distance  $\varepsilon$  satisfies  $0 \leq \varepsilon < \log\left(\frac{1}{p(z)}\right)$  for all  $z \in A$ .

**Lemma 1** and **Assumption 1** give rise to an interpretation of the choice of  $\varepsilon$ , which depends on the baseline distribution of  $S(z, Y)$ . That is, when  $\varepsilon = 0$ , one is completely confident in the baseline probability, while under case (ii) of **Assumption 1**,  $\varepsilon = \log\left(\frac{1}{p(z)}\right)$  corresponds to maximal uncertainty. Note that if  $p(z) = 0$ , then there is no upper limit on the tolerance distance and  $\varepsilon \in [0, \infty)$ , with maximal uncertainty corresponding to  $\varepsilon = \infty$ .

The next result shows that for  $Y$  with positive probability mass at its essential supremum, if the tolerance distance of the uncertainty set is too large, i.e.  $\varepsilon \geq \log\left(\frac{1}{p(z)}\right)$ , then the worst-case probability measure puts all probability mass on the essential supremum of  $Y$ . Thus if  $\varepsilon \geq \log\left(\frac{1}{p(z)}\right)$ , the REF defined in (6) is degenerate justifying **Assumption 1** (ii).

**Proposition 1.** Let  $p(z) > 0$  and be given as in **Assumption 1**. If  $\varepsilon \geq \log\left(\frac{1}{p(z)}\right)$ , then the optimal probability measure  $\mathbb{Q}^\dagger$  attaining the inner supremum in (6) has Radon–Nikodym derivative

$$\frac{d\mathbb{Q}^\dagger}{d\mathbb{P}} = \frac{\mathbb{1}_{\{S(z, Y) = \operatorname{ess sup} S(z, Y)\}}}{p(z)}.$$

**Proof.** First,  $\mathbb{Q}^\dagger$  lies within the KL uncertainty set, since

$$D_{KL}(\mathbb{Q}^\dagger \parallel \mathbb{P}) = \mathbb{E} \left[ \frac{\mathbb{1}_{\{S(z, Y) = \operatorname{ess sup} S(z, Y)\}}}{p(z)} \log \left( \frac{\mathbb{1}_{\{S(z, Y) = \operatorname{ess sup} S(z, Y)\}}}{p(z)} \right) \right] = \log\left(\frac{1}{p(z)}\right).$$

Next,  $\mathbb{E}^{\mathbb{Q}^\dagger}[S(z, Y)] = \operatorname{ess sup} S(z, Y)$  and moreover, for any probability measure  $\mathbb{Q}$  that is absolutely continuous with respect to  $\mathbb{P}$ , it holds that  $\mathbb{E}^{\mathbb{Q}}[S(z, Y)] \leq \operatorname{ess sup} S(z, Y)$ . Hence,  $\mathbb{Q}^\dagger$  attains the inner supremum in (6).  $\square$

In the case when the tolerance distance satisfies [Assumption 1](#), the probability measure attaining the inner supremum in (6) becomes non-degenerate and the REF admits an alternative representation discussed next.

**Theorem 1 (Kullback–Leibler Uncertainty).** Let  $S : \mathcal{A} \times \mathbb{R} \rightarrow [0, \infty)$  be a strictly consistent scoring function for  $R$ , and  $\varepsilon$  such that [Assumption 1](#) is satisfied. Then, the REF has representation

$$\bar{R}^S(Y) = \operatorname{argmin}_{z \in \mathcal{A}} \frac{\mathbb{E}[S(z, Y)e^{\eta^*(z)S(z, Y)}]}{\mathbb{E}[e^{\eta^*(z)S(z, Y)}]},$$

where for each  $z \in \mathcal{A}$ ,  $\eta^*(z) \geq 0$  is the unique solution to  $\varepsilon = D_{KL}(\mathbb{Q}^{\eta(z)} \parallel \mathbb{P})$ , with

$$\frac{d\mathbb{Q}^{\eta(z)}}{d\mathbb{P}} := \frac{e^{\eta(z)S(z, Y)}}{\mathbb{E}[e^{\eta(z)S(z, Y)}]}.$$

**Proof.** Let  $Y$  have cdf  $F$  under  $\mathbb{P}$  and denote its (not-necessarily continuous) probability density function (pdf) by  $f$ , such that e.g.  $\mathbb{E}[Y] = \int yf(y)\nu(dy)$ , where integration is tacitly assumed to be over the support of  $Y$ , which we denote by  $\text{supp}(Y)$ , and where  $\nu$  is either the counting or the Lebesgue measure, depending whether  $Y$  is a continuous or discrete rv. The arguments in this proof can be straightforwardly generalised to  $Y$  being continuously distributed except for a countable number of points, i.e.  $\mathbb{E}[Y] = \int yf(y)dy + \sum_{k \in \mathcal{I}} y_k p_k$ , where  $\mathcal{I} \subseteq \mathbb{N}$  and  $\mathbb{P}(Y = y_k) = p_k > 0$ ,  $k \in \mathcal{I}$ . For simplicity we only work with the representation  $\mathbb{E}[Y] = \int yf(y)\nu(dy)$ .

The inner problem of (6) can be written as an optimisation problem over densities as follows

$$\begin{aligned} \sup_{g: \mathbb{R} \rightarrow \mathbb{R}} \int S(z, y)g(y)\nu(dy), \quad \text{subject to} \quad & \int \frac{g(y)}{f(y)} \log \left( \frac{g(y)}{f(y)} \right) \\ & \times f(y)\nu(dy) \leq \varepsilon, \\ & \int g(y)\nu(dy) = 1, \quad \text{and} \\ & g(y) \geq 0, \\ & \text{for all } y \in \text{supp}(Y). \end{aligned}$$

The above optimisation problem admits the Lagrangian with Lagrange parameters  $\eta_1, \eta_2 \geq 0$ , and  $\eta_3(y) \geq 0$ , for all  $y \in \text{supp}(Y)$ , as

$$\begin{aligned} L(\eta_1, \eta_2, \eta_3, g) = \int & \left( -S(z, y)g(y) + \eta_1 \log \left( \frac{g(y)}{f(y)} \right) g(y) + \eta_2 g(y) \right. \\ & \left. - \eta_3(y)g(y) \right) \nu(dy) - \eta_1 \varepsilon - \eta_2. \end{aligned}$$

The Lagrange parameter  $\eta_3(y)$  guarantees, for all  $y \in \text{supp}(Y)$ , the non-negativity of  $g(\cdot)$ , that is  $g(y) \geq 0$ , whenever  $f(y) > 0$ , and  $g(y) = 0$  otherwise. The associated Euler–Lagrange equation becomes

$$-S(z, y) + \eta_1 \left( \log \left( \frac{g(y)}{f(y)} \right) + 1 \right) + \eta_2 + \eta_3(y) = 0.$$

This implies that

$$\frac{g(y)}{f(y)} = \exp \left\{ \frac{1}{\eta_1} (S(z, y) - \eta_2 - \eta_3(y)) - 1 \right\}.$$

Thus,  $\eta_3(y) \equiv 0$  and imposing  $\eta_2$  yields

$$\frac{g(y)}{f(y)} = \frac{\exp \left\{ \frac{1}{\eta_1} S(z, y) - 1 \right\}}{\int \exp \left\{ \frac{1}{\eta_1} S(z, y) - 1 \right\} f(y)\nu(dy)}.$$

Reparametrising  $\eta := \frac{1}{\eta_1}$ , we define the Radon–Nikodym derivative

$$\frac{d\mathbb{Q}^\eta}{d\mathbb{P}} := \frac{g(Y)}{f(Y)} = \frac{e^{\eta S(z, Y)}}{\mathbb{E}[e^{\eta S(z, Y)}]},$$

which implies that  $Y$  under  $\mathbb{Q}^\eta$  has density  $g$ .

Next, we show that for each  $z \in \mathcal{A}$ , the optimal Lagrangian parameter  $\eta^*$  is the solution to  $\varepsilon = D_{KL}(\mathbb{Q}^{\eta^*} \parallel \mathbb{P})$ , i.e. that KL divergence

constraint is binding. For this we first show that for fixed  $z \in \mathcal{A}$ ,  $D_{KL}(\mathbb{Q}^\eta \parallel \mathbb{P})$  is strictly increasing in  $\eta$ . We calculate

$$\begin{aligned} D_{KL}(\mathbb{Q}^\eta \parallel \mathbb{P}) &= \mathbb{E} \left[ \frac{e^{\eta S(z, Y)}}{\mathbb{E}[e^{\eta S(z, Y)}]} \log \left( \frac{e^{\eta S(z, Y)}}{\mathbb{E}[e^{\eta S(z, Y)}]} \right) \right] \\ &= \mathbb{E} \left[ \frac{e^{\eta S(z, Y)}}{\mathbb{E}[e^{\eta S(z, Y)}]} (\eta S(z, Y) - \log(\mathbb{E}[e^{\eta S(z, Y)}])) \right] \\ &= \eta \frac{\mathbb{E}[e^{\eta S(z, Y)} S(z, Y)]}{\mathbb{E}[e^{\eta S(z, Y)}]} - \log(\mathbb{E}[e^{\eta S(z, Y)}]) \\ &= \eta K'_{S(z, Y)}(\eta) - K_{S(z, Y)}(\eta) =: d(\eta), \end{aligned} \tag{8}$$

where  $K_{S(z, Y)}(\eta) := \log(\mathbb{E}[e^{\eta S(z, Y)}])$ , and  $K'_{S(z, Y)}(\eta) := \frac{\partial}{\partial \eta} K_{S(z, Y)}(\eta)$  denotes the derivative with respect to  $\eta$ . The interchange of the differential operator and expectation is valid as we assume that  $Y$  is essentially bounded. Observe that  $K_{S(z, Y)}(\cdot)$  is the cdf of  $S(z, Y)$ , thus it is differentiable and strictly convex. Therefore  $d(\eta)$  is increasing since

$$d'(\eta) = K'_{S(z, Y)}(\eta) + \eta K''_{S(z, Y)}(\eta) - K'_{S(z, Y)}(\eta) = \eta K''_{S(z, Y)}(\eta) > 0.$$

Furthermore, as the objective function of the inner problem equals the derivative of the cdf, it is also increasing in  $\eta$ . Indeed,

$$\frac{\mathbb{E}[S(z, Y)e^{\eta S(z, Y)}]}{\mathbb{E}[e^{\eta S(z, Y)}]} = K'_{S(z, Y)}(\eta),$$

which is increasing in  $\eta$ . Thus, both the objective function of the inner problem and the KL divergence are strictly increasing in  $\eta$ , the constraint is binding and  $\eta^*$  is the unique solution to  $\varepsilon = D_{KL}(\mathbb{Q}^{\eta^*} \parallel \mathbb{P})$ . To see that a solution to  $\varepsilon = D_{KL}(\mathbb{Q}^\eta \parallel \mathbb{P})$  exists, note that by [Lemma 1](#), it holds that  $\lim_{\eta \rightarrow 0} D_{KL}(\mathbb{Q}^\eta \parallel \mathbb{P}) = 0$  and  $\lim_{\eta \rightarrow +\infty} D_{KL}(\mathbb{Q}^\eta \parallel \mathbb{P}) = \log(\frac{1}{p(z)})$ .

As  $\eta$  and  $\mathbb{Q}^\eta$  depends on  $z$ , we make this dependence explicit in the statement, and write  $\eta(z)$  and  $\mathbb{Q}^{\eta(z)}$ .  $\square$

**Remark 1.** [Theorem 1](#) assumes that  $Y \in \mathcal{L}^\infty$ , though this is an assumption that can be relaxed. Indeed, from the proof of [Theorem 1](#), we see that we only require the mgf of  $S(t, Y)$  at  $\eta^*(t)$ , i.e.,  $\mathbb{E}[e^{\eta^*(t)S(t, Y)}]$ , to be finite for all  $t$  in the neighbourhood of the optimal  $z^*$ . Moreover, there are choice of pairs  $(S, Y)$  of scoring functions and rv for which the REF exists and where  $Y$  is not essentially bounded. An example is  $Y \sim N(\mu, \sigma)$  normally distributed with mean  $\mu$  and standard deviation  $\sigma$ , and the pinball loss,  $S(z, y) = (\mathbb{1}_{\{y \leq z\}} - \alpha)(z - y)$ ,  $\alpha \in (0, 1)$ . Clearly,  $Y$  is not essentially bounded, however, the mgf of the scoring function is finite. To see this, for  $t \in \mathbb{R}$ , we have  $\mathbb{E}[e^{tS(z, Y)}] = \mathbb{E}[e^{t(1-\alpha)(z-Y)^+ + \alpha(z-Y)^-}] \leq \mathbb{E}[e^{t(1-\alpha)(z-Y) + t\alpha(z-Y)}] = \mathbb{E}[e^{t(z-Y)}] = e^{tz} M_Y(-t)$  where  $(\cdot)^+$  and  $(\cdot)^-$  denote the positive and negative part respectively, and  $M_Y(\cdot)$  the mgf of  $Y$ . As the mgf of  $Y$  equals  $M_Y(t) = e^{\mu + \frac{1}{2}\sigma^2 t^2}$ , it is finite for all  $t \in \mathbb{R}$  and [Theorem 1](#) applies.

## 2.2. Existence and uniqueness

For existence and uniqueness of the REF we require additional properties on the scoring functions. Thus, we first recall several definitions related to scoring functions. The first set of properties rely on the first order condition of elicitable functionals. Indeed, as elicitable functionals are defined via an argmin, they can often be found by solving the corresponding first order condition. The following definition of identification functions make this precise. We refer the interested reader to [Steinwart et al. \(2014\)](#) for details and discussions.

**Definition 3 (Identification Function).** Let  $R$  be an elicitable functional. Then, a measurable function  $V : \mathcal{A} \times \mathbb{R} \rightarrow \mathbb{R}$  is

(i) called an **identification function** for  $R$ , if

$$\mathbb{E}[V(z, Y)] = 0 \quad \text{if and only if } z \in R(Y),$$

for all  $z \in \text{Im}^o R$ , where  $\text{Im}^o R$  is the interior of  $\text{Im} R$ , and for all  $Y \in \mathcal{L}^\infty$ .

(ii) **orientated** if  $V$  is an identification function and

$$\mathbb{E}[V(z, Y)] > 0 \text{ if and only if } z > R(Y)^u,$$

for all  $z \in \text{Im}^o R$ , and for all  $Y \in \mathcal{L}^\infty$ .

(iii) The functional  $R$  is called **identifiable** if there exists an identification function for  $R$ .

An identification function thus characterises the first order condition of elicitable functionals and is therefore intimately connected to scoring functions. Moreover, an oriented identification function can further rank predictions. To see this, note that any identification function gives raise to a scoring function, and in particular, any oriented identification function gives raise to an order sensitive scoring function, see e.g. Steinwart et al. (2014). A scoring function is called **order sensitive** (or accuracy rewarding) for  $R$ , if for all  $t_1, t_2 \in \mathbf{A}$  with either  $R(Y)^u < t_1 < t_2$  or  $R(Y)^l > t_1 > t_2$  it holds that

$$\mathbb{E}[S(R(Y), Y)] < \mathbb{E}[S(t_1, Y)] < \mathbb{E}[S(t_2, Y)].$$

Thus, the further away the prediction  $t_2$  is from  $R(Y)$ , the larger its expected score.

Moreover, recall that a scoring function  $S$  is locally Lipschitz continuous in  $z$ , if for all intervals  $[a, b] \subseteq \mathbf{A}$ , there exists a constant  $c_{a,b} \geq 0$  such that for all  $z_1, z_2 \in [a, b]$  and all  $y \in \text{supp}(Y)$ , we have

$$|S(z_1, y) - S(z_2, y)| \leq c_{a,b} |z_1 - z_2|.$$

We next establish the following result, which characterises identification functions of exponentially transformed scoring functions, that will be instrumental in proving existence of the REF.

**Lemma 2.** Let  $S : \mathbf{A} \times \mathbb{R} \rightarrow [0, \infty)$  be strictly convex in and locally Lipschitz continuous in its first component, and a strictly consistent scoring function for a functional  $R$ . Moreover, define  $H(z, y) := e^{vS(z, y)}$ , for  $v > 0$ .

Then,  $H(\cdot, \cdot)$  is strictly convex in  $z$  and locally Lipschitz continuous in  $z$ . Moreover,  $R^H := \underset{z \in \mathbf{A}}{\text{argmin}} \mathbb{E}[H(z, Y)]$  exists for  $\mathbf{A} \subseteq \mathbb{R}$ , thus  $R^H$  is elicitable with scoring function  $H(\cdot, \cdot)$ . Furthermore, there exists an oriented identification function  $W$  for  $R^H$  that satisfies

$$k(z) W(z, y) = v e^{vS(z, y)} \frac{\partial}{\partial z} S(z, y), \quad \text{for almost all}$$

$$(z, y) \in \text{Im}^o R \times \text{supp}(Y),$$

and for some  $k(z) > 0$ .

**Proof.** Clearly,  $H$  is strictly convex in  $z$  as it is a composition of the monotone and convex exponential function and the scoring function, which is convex in  $z$ . Similarly, it is locally Lipschitz continuous in  $z$  as  $H$  is the composition of two locally Lipschitz functions. As  $H$  is strictly convex in  $z$ , we have that  $\mathbb{E}[H(z, Y)]$  is also convex in  $z$  for all  $Y \in \mathcal{L}^\infty$ . Thus,  $R^H$  exists and is an elicitable functional.

By Corollary 9 of Steinwart et al. (2014),  $R^H$  is identifiable and has an oriented identification function, which we denote by  $W$ . By (iii), Theorem 8 of Steinwart et al. (2014), the identification function satisfies for almost all  $(z, y)$

$$k(z) W(z, y) = \frac{\partial}{\partial z} H(z, y) = v e^{vS(z, y)} \frac{\partial}{\partial z} S(z, y),$$

for some  $k(z) > 0$ ,  $z \in \mathbf{A}$ , and where the last equation holds by definition of  $H$ .  $\square$

The next results shows the conditions for the REF to exist, and when it is unique.

**Theorem 2 (Existence and Uniqueness).** Let  $R$  be an elicitable functional with strictly consistent scoring function  $S$ , and  $\epsilon$  such that Assumption 1 is satisfied. Further assume that  $S(z, y)$  is strictly convex in  $z$ , and continuously differentiable in  $z$ . Assume that  $\int_{\mathbf{A}} |\frac{\partial}{\partial z} S(z, y)| dz < \infty$ , for all  $y \in \text{supp}(Y)$ . Then the following hold:

- (i) there exists a solution to optimisation problem (6), that is  $\bar{R}^S(Y)$  exists,
- (ii) if  $\underset{z \in \mathbf{A}}{\text{argmin}} \mathbb{E}[e^{vS(z, Y)}]$  is a singleton for all  $v > 0$ , then the solution to optimisation problem (6) is unique.

**Proof.** We define the value function of the inner optimisation problem of (6) by

$$J(z) := \sup_{Q \in \mathcal{Q}_\epsilon} \mathbb{E}^Q[S(z, Y)]. \quad (9)$$

For existence (i) we first apply the envelope theorem for saddle point problems, Theorem 4 in Milgrom and Segal (2002) to derive an expression for  $\frac{d}{dz} J(z)$  and second show that  $\frac{d}{dz} J(z)$  crosses zero. For (ii) we show that  $\frac{d}{dz} J(z)$  crosses zero at most once, thus  $J(z)$  admits a unique minima.

**Part 1:** Rewriting optimisation problem (9) as a constrained optimisation problem over densities gives

$$\sup_{\text{density } g} \int S(z, y) g(y) v(dy), \quad \text{subject to } \int g(y) \log \left( \frac{g(y)}{f(y)} \right) v(dy) \leq \epsilon.$$

As the space of densities is convex and as the objective function  $\mathcal{J}(g, z) := \int S(z, y) g(y) v(dy)$  and the constraint function  $c(g, z) := \epsilon - \int_R g(y) \log \left( \frac{g(y)}{f(y)} \right) dy$  are both concave in  $g$ , the constrained optimisation problem can be represented as a saddle-point problem with associated Lagrangian

$$L(g, \eta, z) := \mathcal{J}(g, z) + \eta c(g, z).$$

Moreover, it holds that

$$J(z) = L(g^*(z), \eta^*(z), z),$$

for saddle points  $(g^*(z), \eta^*(z))$ . Next, we apply Theorem 4 in Milgrom and Segal (2002) to the Lagrangian  $L(g, \eta, z)$ . For this note that for fixed  $(g, \eta)$ ,  $L(g, \eta, \cdot)$  is absolutely continuous, since the scoring function is continuously differentiable in  $z$ . Moreover, for each  $z$ , the set of saddle points is non-empty by Theorem 1. Also, there exists a non-negative and integrable function  $b : \mathbf{A} \rightarrow [0, \infty)$  such that

$$\left| \frac{\partial}{\partial z} L(g, \eta, z) \right| = \left| \int \frac{\partial}{\partial z} S(z, y) g(y) v(dy) \right| \leq b(z),$$

where the inequality follows since the scoring function is locally Lipschitz continuous in  $z$  (since it is continuously differentiable), and  $g$  is a density. Integrability of  $b$  follows by the integrability assumption on the derivative of the scoring function. Also,  $\frac{\partial}{\partial z} L(g, \eta, z) = \frac{\partial}{\partial z} \mathcal{J}(g, z)$  is continuous in  $g$  and  $L(g, \eta, z)$  equi-differentiable in  $g$  and  $\eta$ . Thus the assumptions of Theorem 4 in Milgrom and Segal (2002) are satisfied and it holds that

$$J(z) = J(0) + \int_0^z \frac{\partial}{\partial s'} L(g, \eta, s') \Big|_{g=g^*(s), \eta=\eta^*(s), s'=s} ds. \quad (10)$$

Therefore, taking derivative with respect to  $z$  of (10), we have

$$\begin{aligned} \frac{d}{dz} J(z) &= \frac{\partial}{\partial z} L(g, \eta, z) \Big|_{g=g^*(z), \eta=\eta^*(z)} \\ &= \int \frac{\partial}{\partial z} S(z, y) g(y) v(dy) \Big|_{g=g^*(z), \eta=\eta^*(z)} \\ &= \frac{\mathbb{E} \left[ \left\{ \frac{\partial}{\partial z} S(z, Y) \right\} e^{\eta^*(z) S(z, Y)} \right]}{\mathbb{E} [e^{\eta^*(z) S(z, Y)}]}, \end{aligned}$$

where in the last equality we used that  $g^*(z)$  and  $\eta^*(z)$  are given in Theorem 1.

**Part 2:** To show that  $\frac{d}{dz} J(z)$  crosses zero, we proceed as follows. For fixed  $\eta > 0$ , define  $H^\eta(z, y) := e^{\eta S(z, y)}$  and  $\bar{H}^\eta(z) := \mathbb{E}[H^\eta(z, Y)]$ . By Lemma 2,  $H^\eta(\cdot, \cdot)$  is strictly convex in  $z$ , locally Lipschitz continuous, and  $[z_\eta^u, z_\eta^u] := \underset{z \in \mathbf{A}}{\text{argmin}} \bar{H}^\eta(z)$  exists. Furthermore, by Lemma 2, there exists an oriented identification function for  $z_\eta^*$  that satisfies

$$k(z) W^\eta(z, y) = \eta e^{\eta S(z, y)} \frac{\partial}{\partial z} S(z, y),$$

for some  $k(z) > 0$ , and for all  $z \in A$ . Since for each  $\eta > 0$ ,  $W^\eta$  is an oriented identification function and as  $k(\cdot) \geq 0$ , we have that for all  $z > z_\eta^u$ ,

$$k(z)\mathbb{E}[W(z, Y)] = \mathbb{E}[e^{\eta S(z, Y)} \frac{\partial}{\partial z} S(z, Y)] > 0,$$

and similarly for all  $z < z_\eta^l$ ,

$$k(z)\mathbb{E}[W(z, Y)] = \mathbb{E}[e^{\eta S(z, Y)} \frac{\partial}{\partial z} S(z, Y)] < 0.$$

Since this holds for all  $\eta > 0$ , the above equations also hold for  $\eta^*$ . Therefore

$$\begin{aligned} \frac{d}{dz} J(z) &> 0, \quad \text{for all } z > z_{\eta^*}^u, \\ \frac{d}{dz} J(z) &< 0, \quad \text{for all } z < z_{\eta^*}^l, \quad \text{and} \\ \frac{d}{dz} J(z) \Big|_{z \in [z_{\eta^*}^l, z_{\eta^*}^u]} &= 0. \end{aligned}$$

Therefore,  $J(z)$  admits a (potentially interval valued) minima.

To show (ii), assume that  $\operatorname{argmin}_{z \in A} \mathbb{E}[e^{\eta S(z, Y)}]$  is a singleton for all  $v > 0$ . That is  $z_{\eta^*}^l = z_{\eta^*}^u$  and is a singleton for all  $\eta > 0$ . Which implies that  $J(z)$  admits a unique minima.  $\square$

We are able to alternatively write the REF in terms of the cgf of the scoring function. For this, we denote the derivative of the cgf of  $Y \in \mathcal{L}^\infty$  by  $K'_Y(t) := \frac{\partial}{\partial t} K_Y(t)$ .

**Proposition 2 (Alternative Representation).** *The REF can be represented as:*

$$T^{KL}(Y) = \operatorname{argmin}_{z \in A} K'_{S(z, Y)}(\eta^*(z)),$$

where for each  $z \in A$ ,  $\eta^*(z)$  is the unique solution to

$$\eta(z) K'_{S(z, Y)}(\eta(z)) - K_{S(z, Y)}(\eta(z)) = \varepsilon. \quad (11)$$

**Proof.** Note that by definition, the cgf of  $S(z, Y)$  is  $K_{S(z, Y)(\eta)} = \log(\mathbb{E}[e^{\eta S(z, Y)}])$ . Then, we have

$$K'_{S(z, Y)}(\eta^*(z)) = \frac{\mathbb{E}[S(z, Y)e^{\eta^*(z)S(z, Y)}]}{\mathbb{E}[e^{\eta^*(z)S(z, Y)}]}.$$

Moreover, Eq. (11) for  $\eta(z)$  is equivalent to Eq. (8). Thus, the two optimisation problems are equivalent.  $\square$

### 3. Choice of scoring function

As elicitable functionals are defined as the  $\operatorname{argmin}$  of an expected scoring function, the scoring function is not unique, indeed, there are infinitely many scoring functions. In this section, we discuss the family of homogeneous scoring function that naturally lead to illustration of the REF via Murphy diagrams.

#### 3.1. Families of $b$ -homogeneous scoring functions

To investigate the effect of the scoring function on the REF, we propose the use of  $b$ -homogeneous scoring functions as argued in Efron (1991) and Patton (2011), and studied in Nolde and Ziegel (2017).

**Definition 4 ( $b$ -Homogeneous Scores).** A scoring function  $S : A' \times \mathbb{R} \rightarrow [0, \infty)$ ,  $A' \subseteq \mathbb{R}$ , is positively homogeneous of degree  $b \in \mathbb{R}$ , if

$$S(cz, cy) = c^b S(z, y)$$

for all  $(z, y) \in A' \times \mathbb{R}$  and for all  $c > 0$ .

We say that a scoring function is positively homogeneous if there exists a  $b \in \mathbb{R}$  such that it is positively homogeneous of degree  $b$ .

These families of parameterised scoring functions retain useful properties of the elicitable functional, discussed next.

**Proposition 3.** Let  $S$  be a strictly consistent scoring function for  $R$ , then the following holds:

- (i) If  $S$  is a positive homogeneous scoring function, then  $R$  and  $\bar{R}^S$  are positive homogeneous of degree 1 for  $A' \in \{[0, \infty), (-\infty, 0], \mathbb{R}\}$ .
- (ii) If  $A' = \mathbb{R}$  and  $S(z - c, y) = S(z, y + c)$  for  $c \in \mathbb{R}$ , then  $R$  and  $\bar{R}^S$  are translation invariant.
- (iii) If  $S(y, y) = 0$  and  $y \in A'$ , then  $R(m) = m$  and  $\bar{R}^S(m) = m$  for all  $m \in \mathbb{R}$ . In particular,  $R(0) = 0$  and  $\bar{R}^S(0) = 0$ .

**Proof.**

(i) The case for  $R$  follows from Fissler and Pesenti (2023). Let  $A' \in \{[0, \infty), (-\infty, 0], \mathbb{R}\}$ . To show that it also holds for  $\bar{R}^S$ , let  $S$  be a positive homogeneous scoring function for  $R$  of degree  $b$ . Then, using the change of variable  $z := cw$ , we obtain

$$\begin{aligned} \bar{R}^S(cY) &= \operatorname{argmin}_{z \in A'} \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q [S(z, cY)] \\ &= c \operatorname{argmin}_{w \in A'} \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q [S(cw, cY)] \\ &= c \operatorname{argmin}_{w \in A'} \sup_{Q \in \mathcal{Q}} c^b \mathbb{E}^Q [S(w, Y)] \\ &= c \operatorname{argmin}_{w \in A'} \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q [S(w, Y)] \\ &= c \bar{R}^S(Y). \end{aligned}$$

(ii) Suppose that  $S(z - c, y) = S(z, y + c)$  for  $c \in \mathbb{R}$ . Fix  $c \in \mathbb{R}$ , then

$$\begin{aligned} \bar{R}^S(Y + c) &= \operatorname{argmin}_{z \in \mathbb{R}} \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q [S(z, Y + c)] \\ &= \operatorname{argmin}_{z \in \mathbb{R}} \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q [S(z - c, Y)] \\ &= \operatorname{argmin}_{z \in \mathbb{R}} \left\{ \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q [S(z, Y)] \right\} + c \\ &= \bar{R}^S(Y) + c. \end{aligned}$$

The statement for  $R$  follows by setting  $\varepsilon = 0$ .

(iii) Assume that  $S(y, y) = 0$  for  $y \in A'$  and let  $m \in \mathbb{R}$ . Then

$$\bar{R}^S(m) = \operatorname{argmin}_{z \in A'} \sup_{Q \in \mathcal{Q}} \mathbb{E}^Q [S(z, m)] = \operatorname{argmin}_{z \in A'} S(z, m) = m.$$

The result for  $R$  follows for the special case of  $\varepsilon = 0$ .  $\square$

In the risk management literature, significant emphasis is placed on what is considered desirable properties of a risk measure. Among these properties, positive homogeneity and translation invariance enjoy significant interest due to their inclusion into the notion of a coherent risk measure (Artzner et al., 1999). Here, we show that we can translate these properties on the elicitable functional (risk measure) into properties on the corresponding scoring function, and that there is a relationship between these properties on the scoring function to their elicitable functional.

In particular, the positive homogeneity property, Proposition 3 case (i), allows for the rescaling of the rvs. This has the financial interpretation of allowing for currency and unit conversions. This also has practical implications, in that a practitioner can rescale  $Y$ , then estimate REF, for example to improve numerical performance, and then rescale the estimator back to the original magnitude, all using Proposition 3 case (i). Translation invariance, or cash invariance, is typically motivated by the risk-free property of cash assets. Meaning that by adding a constant value the risk of the random portfolio should be reduced by the same amount. A property that may be preserved by the REF.

#### 3.2. Murphy diagrams for robust elicitable functionals

In this section we illustrate the REF on different functionals such as the mean, VaR, and expectiles using  $b$ -homogeneous scoring functions.

The  $b$ -homogeneous class of scoring functions allows practitioners to rescale losses via the homogeneity property of the REF, to improve numerical stability of the functional.

**Proposition 4** ( *$b$ -homogeneous Scoring Functions — mean* (Nolde & Ziegel, 2017)). *The class of strictly consistent and  $b$ -homogeneous scoring functions  $S_b^{\mathbb{E}} : [0, \infty)^2 \rightarrow \mathbb{R}$  for the mean satisfying  $S(y, y) = 0$  are given by any positive multiple of a member of the Patton family*

$$S_b^{\mathbb{E}}(z, y) = \begin{cases} \frac{y^b - z^b}{b(b-1)} - \frac{z^{b-1}}{b-1}(y-z), & b \in \mathbb{R} \setminus \{0, 1\}, \\ \frac{y}{z} - \log\left(\frac{y}{z}\right) - 1, & b = 0, \\ y \log\left(\frac{y}{z}\right) - (y-z), & b = 1, \end{cases} \quad (12)$$

where we require that  $z, y > 0$ .

Note that the squared-loss,  $S(z, y) = (z-y)^2$ , is recovered when  $b = 2$  in the  $b$ -homogeneous scoring function for the mean in Eq. (12).

Introduced in Ehm et al. (2016), a Murphy diagram is a graph used to display the effect of a scoring function's homogeneity parameter against the value of the functional. Here, we use this idea to plot the function  $b \mapsto \bar{R}^{S_b}$ . In the next examples, we plot the REF Murphy diagrams for the mean, VaR, and expectiles.

For the numerical examples, and as we work on  $\mathcal{L}^\infty$ , we consider right truncated rvs. Right truncated rvs arise in financial and insurance contexts via financial options, limits on (re)insurance contracts, and maximal losses, e.g. if an insured asset is written off as a total loss. We also refer to Albrecher et al. (2017) for use of truncated and censored distributions in reinsurance settings. Let  $X$  be a random variable with pdf  $g$  and cdf  $G$ . Then the right truncated random variable  $Y := X | X \leq \bar{x}$  with truncation point  $\bar{x} \in \mathbb{R}$ , has pdf

$$f_Y(y | X \leq \bar{x}) = \frac{g(y)}{G(\bar{x})} \mathbb{1}_{\{y \leq \bar{x}\}}.$$

Furthermore, it holds that  $F^{-1}(\alpha) = G^{-1}(\alpha G(\bar{x}))$ , and  $\mathbb{E}[Y] = \frac{\int_0^{\bar{x}} x g(x) dx}{F(\bar{x})}$ , whenever  $Y \geq 0$   $\mathbb{P}$ -a.s.

In the examples below, we consider right truncated exponential losses. In particular, we truncate each exponential to its 95% quantile i.e. we set the truncation point to  $\bar{x} := F^{-1}(0.95)$ , where  $F^{-1}$  is the quantile function of the exponential distribution. This corresponds to retaining 95% of the support of the exponential distribution. We denote the above described distribution by  $\text{TExp}(\lambda)$ , where  $\lambda$  is the parameter of the original exponential distribution.

**Example 1** (*Murphy Diagrams for the Mean*). We consider the mean functional and its REF, the REF mean, with the  $b$ -homogeneous scoring functions given in Proposition 4 for the Beta distribution. Here we use the convention that  $\text{Beta}(\beta_1, \beta_2)$  has density  $f(x) = \frac{\Gamma(\beta_1 + \beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} x^{\beta_1-1} (1-x)^{\beta_2-1}$ , where  $\Gamma(\cdot)$  is the Gamma function. Fig. 1 displays the REF mean with varying uncertainty tolerances  $\varepsilon$  between 0 and 0.5, for the  $\text{Beta}(\beta_1 = 2, \beta_2 = 2)$  baseline distribution in the left panel of the figure and the  $\text{TExp}(2)$  baseline distribution in the right panel. We observe that for the Beta distribution, for each  $\varepsilon$ , the REF mean is increasing in the homogeneity degree  $b$  and converges to the value 0.5, that is the mean of the Beta distribution. Furthermore, for fixed  $b$ , the larger the  $\varepsilon$ , the smaller the REF mean. In particular, the REF mean is always less than the baseline mean. This is not observed in the right panel, which displays the REF mean for the  $\text{TExp}(2)$  distribution, where the REF mean is always greater than the baseline mean of 0.4. Moreover, for each  $\varepsilon$  the REF mean is increasing in the homogeneity degree  $b$ , and for fixed  $b$  the REF mean is ordered in  $\varepsilon$ . Therefore, we observe that depending on the underlying distribution the REFs can be greater or smaller than the baseline mean. Fig. 1 also indicates that there is a trade-off between  $\varepsilon$  and  $b$ , indicating that the choice of tolerance  $\varepsilon$  should be influenced by the homogeneity degree  $b$  of the scoring function. Next, we consider how the REF mean changes according

to the parameters of the Beta distribution. In particular Fig. 2 show the REF mean for fixed  $b = 1.5$  against different values of the shape parameters of the Beta distribution. The left panel displays variation with  $\beta_1$  and the right panel with  $b$ , where the different lines correspond to different  $\varepsilon$ . Interestingly, the REF mean for different  $\varepsilon$  cross when plotted against the first parameter of the Beta distribution, whereas the REF mean are ordered in  $\varepsilon$  for the second shape parameter  $\beta_2$ .

Another elicitable functional of interest is the Value-at-Risk (VaR), also known as the quantile.

**Definition 5** (*Value-at-Risk*). The Value-at-Risk at tolerance level  $\alpha \in (0, 1)$  is defined for  $X \in \mathcal{L}^\infty$ , as

$$\text{VaR}_\alpha(X) = \inf \{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq \alpha\}.$$

The VaR is well-known to be elicitable and its family of  $b$ -homogeneous scoring functions is recalled next.

**Proposition 5** ( *$b$ -homogeneous Scoring Functions — VaR* (Nolde & Ziegel, 2017)). *The class of strictly consistent and  $b$ -homogeneous scoring functions  $S_b^{\text{VaR}} : \mathbb{R}^2 \rightarrow [0, \infty)$  for  $\text{VaR}_\alpha$  satisfying  $S(y, y) = 0$  are given by*

$$S_b^{\text{VaR}}(z, y) = (\mathbb{1}_{\{y \leq z\}} - \alpha)(g(z) - g(y)),$$

where

$$g(y) = \begin{cases} d_1 y^b \mathbb{1}_{\{y > 0\}} - d_2 |y|^b \mathbb{1}_{\{y < 0\}} & \text{if } b > 0 \text{ and } y \in \mathbb{R}, \\ d \log(y) & \text{if } b = 0 \text{ and } y > 0, \\ -d y^b & \text{if } b < 0 \text{ and } y > 0, \end{cases}$$

for positive constants  $d, d_1, d_2 > 0$ .

The  $b$ -homogeneous scoring function for the VaR coincides with the pinball loss scoring function when  $b = 1$ . For simplicity, we choose  $d = d_1 = d_2 = 1$  in the following experiments.

**Example 2** (*Murphy Diagrams for the VaR*). Similarly to Example 1, we plot the Murphy diagrams for the VaR functional for the  $\text{TExp}(2)$  baseline distribution. Fig. 3 displays in the left panel the REF VaRs against the homogeneity degree  $b$  for the baseline distribution  $\text{TExp}(\lambda = 2)$ , on the centre panel against the parameter of the  $\text{TExp}(\lambda)$  for fixed  $b = 1.5$ . The right most panel considers the Beta distribution as the baseline and plots the REF VaR against the second shape parameter,  $\beta_2$  of the Beta distribution. We observe in all three panels that the REF VaRs are ordered in  $\varepsilon$  with larger values for larger uncertainty tolerances  $\varepsilon$ . Note that in the centre panel, for small  $\lambda$ , the difference between the REF VaR and the baseline VaR of the  $\text{TExp}(\lambda)$  is significantly larger than the corresponding difference for large  $\lambda$ s. In the right panel, we observe that the REF VaR can be larger or smaller than the baseline VaR.

Another commonly considered functional is the expectile, which has been proposed for use in risk management in Bellini et al. (2014). Its elicitability is established in Gneiting (2011).

**Definition 6** ( *$\tau$ -expectile*). Let  $\tau \in (0, 1)$ , then the  $\tau$ -expectile,  $e_\tau(\cdot)$ , is the unique solution to

$$\tau \mathbb{E}[(X - e_\tau(X))^+] = (1 - \tau) \mathbb{E}[(X - e_\tau(X))^-].$$

For  $\tau \geq \frac{1}{2}$ , the expectile is coherent in the sense of Artzner et al. (1999). The family of  $b$ -homogeneous scoring functions for the expectile is as follows.

**Proposition 6** ( *$b$ -homogeneous Scoring Functions — expectile* (Nolde & Ziegel, 2017)). *For  $\tau \in (0, 1)$ , the strictly consistent  $b$ -homogeneous scoring function  $S_b^\tau : [0, \infty)^2 \rightarrow \mathbb{R}$  corresponding to the  $\tau$ -expectile is given as*

$$S_b^\tau(z, y) = |1 - \tau| \cdot S_b^{\mathbb{E}}(z, y),$$

where  $S_b^{\mathbb{E}}(z, y)$  is the  $b$ -homogeneous scoring function for the mean as defined in Proposition 4, and where  $z, y > 0$ .

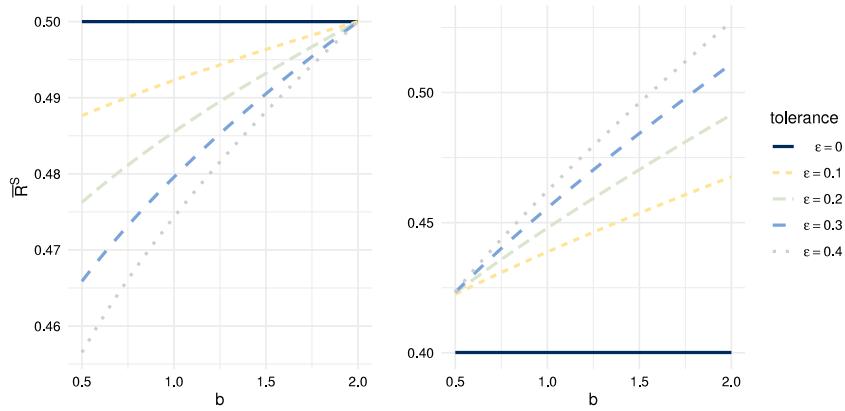


Fig. 1.  $\bar{R}^S$  for varying  $b$  parameter of the  $b$ -homogeneous mean scoring function for Beta(2,2) distribution (left), and TExp(2) (right).

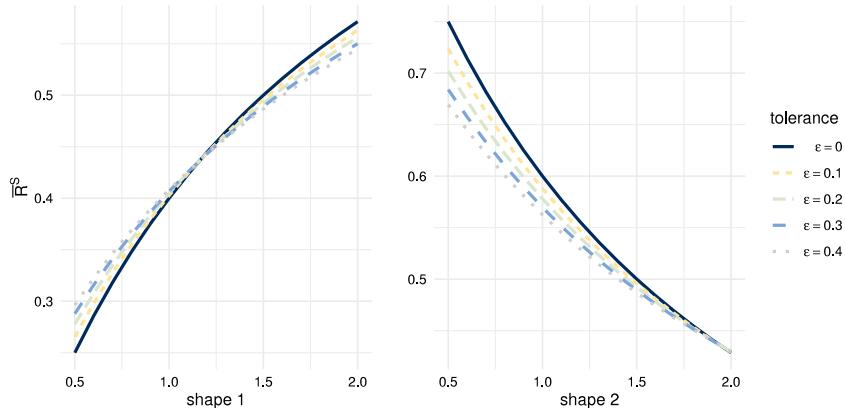


Fig. 2.  $\bar{R}^S$  for varying shape parameters of the Beta distribution with  $b = 1.5$  homogeneous mean scoring function,  $\beta_2 \equiv 1.5$  (left), and  $\beta_2 \equiv 1.5$  (right).

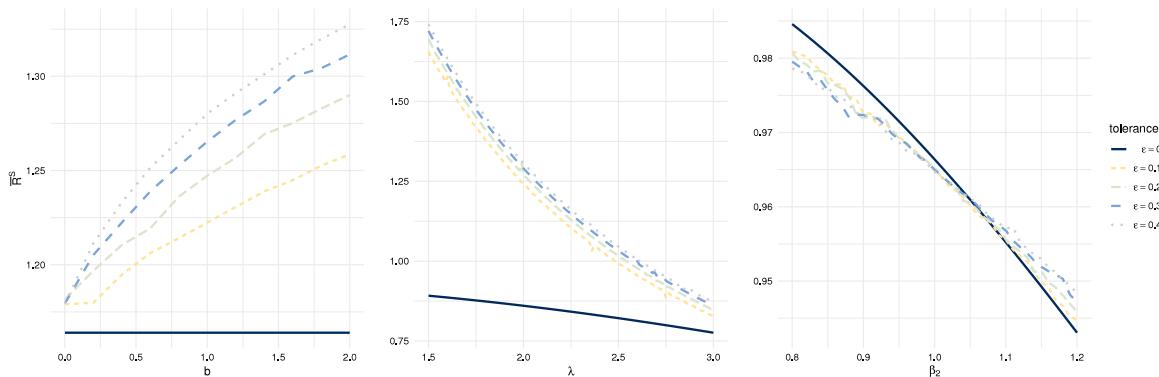


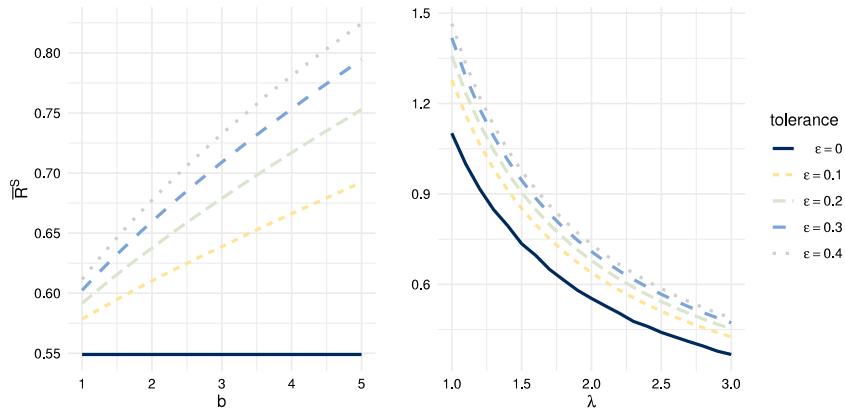
Fig. 3. REF VaR at level  $\alpha = 0.95$ . Left:  $\bar{R}^S$  against the homogeneity parameter  $b$  for baseline distribution TExp(2). Centre:  $\bar{R}^S$  against the parameter  $\lambda$  of the baseline distribution TExp( $\lambda$ ) and fixed  $b = 1.5$ . Right:  $\bar{R}^S$  against  $\beta_2$  of the baseline distribution Beta(1.5,  $\beta_2$ ) and fixed  $b = 1.5$ .

**Example 3 (Murphy Diagrams for the Expectile).** We continue to consider the underlying distribution of the truncated exponential as described in [Example 2](#). We calculate the 0.7-expectile of the baseline TExp(3) and TExp( $\lambda$ ) numerically using a simulated sample of size 30,000.

The left panel of [Fig. 4](#) displays the Murphy diagram of the  $b$ -homogeneous score for the expectile as defined in [Proposition 6](#), and the right panel displays the robust expectile against the TExp parameter  $\lambda$ . The left panel in [Fig. 4](#) exhibits similar behaviour as in previous examples for the mean and VaR. In particular, we have that for fixed  $b$ , the REF is increasing in tolerance  $\epsilon$ . This is again seen in the right panel, where for any fixed  $\lambda$ , the REF is increasing in  $\epsilon$ .

#### 4. Multivariate robust elicability functional

Many statistical functionals are not elicitable on their own, such as the variance, Range-Value-at-Risk, Expected Shortfall (ES), and distortion risk measures, [Fissler and Ziegel \(2021\)](#), [Gneiting \(2011\)](#), [Kou and Peng \(2016\)](#) and [Wang and Ziegel \(2015\)](#). However, they may be elicitable as multi-dimensional functionals, as is in the case of the pairs (mean, variance) and (VaR, ES) ([Fissler et al., 2016](#)). In this section, we generalise the REF to these instances where the functionals are elicitable in the multi-dimensional case.



**Fig. 4.**  $\bar{R}^S$  for varying  $b$  parameter of the  $b$ -homogeneous 0.7-expectile scoring function (left), and varying  $\lambda$  with  $b = 3$  homogeneous 0.7-expectile scoring function (right), TExp(2) distribution.

#### 4.1. The notion of $k$ -dimensional elicibility

In Fissler and Ziegel (2016), the authors establish the framework for  $k$ -consistency and  $k$ -elicability, recalled next.

**Definition 7** ( $k$ -elicability (Fissler & Ziegel, 2016)). Let  $A \subseteq \mathbb{R}^k$ ,  $k \in \mathbb{N}$ , be an action domain and  $\mathbf{R}: \mathcal{M}^\infty \rightarrow A$  be a  $k$ -dimensional law-invariant functional.

(i) A scoring function  $S: A \times \mathbb{R} \rightarrow \mathbb{R}$  is **consistent** for a functional  $\mathbf{R}$ , if

$$\int S(\mathbf{t}, y) dF(y) \leq \int S(\mathbf{z}, y) dF(y), \quad (13)$$

for all  $F \in \mathcal{M}^\infty$ ,  $\mathbf{t} \in \mathbf{R}(F)$ , and all  $\mathbf{z} \in A$ .

(ii) A scoring function  $S$  is **strictly consistent** for a functional  $\mathbf{R}$ , if it is consistent and Eq. (13) holds with equality only if  $\mathbf{z} \in \mathbf{R}(F)$ .

(iii) A functional  $\mathbf{R}$  is  **$k$ -elicitable**, if there exists a strictly consistent scoring function  $S$  for  $\mathbf{R}$ .

Similar to the one-dimensional case, elicitable functionals and strictly consistent scoring functions have a correspondence relationship. Indeed, a  $k$ -elicitable functional  $\mathbf{R}(\cdot) := (R_1(\cdot), \dots, R_k(\cdot))$  admits the representation

$$(R_1(Y), \dots, R_k(Y)) = \operatorname{argmin}_{\mathbf{z} \in A} \int S(\mathbf{z}, y) dF_Y(y),$$

for all  $F_Y \in \mathcal{M}^\infty$  and where  $S$  is any strictly consistent scoring function for  $\mathbf{R}$ .

Similar to univariate functionals, in the multi-dimensional setting, it is of interest to consider uncertainty in the baseline distribution or measure  $\mathbb{P}$ , and we define the multi-dimensional REF as follows. Let  $\mathbf{R}$  be an elicitable functional with strictly consistent scoring function  $S$  and let  $\epsilon \geq 0$ . Then we define the  **$k$ -dimensional robust elicitable functional** (REF) evaluated at  $Y \in \mathcal{L}^\infty$ , by

$$\bar{\mathbf{R}}^S(Y) := \operatorname{argmin}_{\mathbf{z} \in A} \sup_{\mathbb{Q} \in \mathcal{Q}_\epsilon} \mathbb{E}^{\mathbb{Q}}[S(\mathbf{z}, Y)], \quad (14)$$

where the uncertainty set  $\mathcal{Q}_\epsilon$  is given in Eq. (7). As  $\bar{\mathbf{R}}^S$  is  $k$ -dimensional, we write  $\bar{R}_i^S$  for the  $i$ th component of  $\bar{\mathbf{R}}^S$ , thus  $\bar{\mathbf{R}}^S(\cdot) := (\bar{R}_1^S(\cdot), \dots, \bar{R}_k^S(\cdot))$ .

The results of the univariate case Theorem 1 follow readily into the multi-dimensional setting. For this, we first generalise the assumption on the tolerance distance  $\epsilon$ .

**Assumption 2** (Maximal Kullback–Leibler Distance). Denote  $p(\mathbf{z}) := \mathbb{P}(S(\mathbf{z}, Y) = \text{ess sup } S(\mathbf{z}, Y))$ . Then one of the following holds:

(i) If  $p(\mathbf{z}) = 0$  for all  $\mathbf{z} \in A$ , then the tolerance distance satisfies  $\epsilon \in [0, \infty)$ .

(ii) If  $p(\mathbf{z}) > 0$  for some  $\mathbf{z} \in A$ , then the tolerance distance  $\epsilon$  satisfies

$$0 \leq \epsilon < \log\left(\frac{1}{p(\mathbf{z})}\right) \quad \text{for all } \mathbf{z} \in A.$$

Under this condition, Theorem 1 holds in the  $k$ -dimensional setting.

**Corollary 1** (Kullback–Leibler Uncertainty). Let  $\mathbf{R}$  be a  $k$ -elicitable functional,  $S: A \times \mathbb{R} \rightarrow [0, \infty)$  be a strictly consistent scoring function for  $\mathbf{R}$ , and  $\epsilon$  such that Assumption 2 is satisfied. Then, the REF has representation

$$\bar{\mathbf{R}}^S(Y) = \operatorname{argmin}_{\mathbf{z} \in A} \frac{\mathbb{E}[S(\mathbf{z}, Y) e^{\eta^*(\mathbf{z}) S(\mathbf{z}, Y)}]}{\mathbb{E}[e^{\eta^*(\mathbf{z}) S(\mathbf{z}, Y)}]}, \quad (15)$$

where each  $\mathbf{z} \in A$ ,  $\eta^*(\mathbf{z}) \geq 0$  is the unique solution to  $\epsilon = D_{KL}(\mathbb{Q}^{\eta^*(\mathbf{z})} \parallel \mathbb{P})$ , with

$$\frac{d\mathbb{Q}^{\eta^*(\mathbf{z})}}{d\mathbb{P}} := \frac{e^{\eta^*(\mathbf{z}) S(\mathbf{z}, Y)}}{\mathbb{E}[e^{\eta^*(\mathbf{z}) S(\mathbf{z}, Y)}]}.$$

**Proof.** The proof is similar to that of Theorem 1. For simplicity we let  $Y$  be continuously distributed with density  $f(\cdot)$  under  $\mathbb{P}$ . Analogous to Theorem 1, this can be generalised to  $Y$  being discrete or continuously distributed except for a countable number of points.

Then the inner optimisation problem of (15) can be written as an optimisation problem over the density of  $Y$  as follows

$$\begin{aligned} \sup_{g: \mathbb{R} \rightarrow \mathbb{R}} \int S(\mathbf{z}, y) g(y) dy, \quad \text{subject to} \int \frac{g(y)}{f(y)} \log\left(\frac{g(y)}{f(y)}\right) f(y) dy \leq \epsilon, \\ \int g(y) dy = 1, \quad \text{and} \\ g(y) \geq 0, \quad \text{for all } y \in \text{supp}(Y). \end{aligned}$$

This optimisation problem admits the Lagrangian, with Lagrange parameters  $\eta_1, \eta_2 \geq 0$  and  $\eta_3(y) \geq 0$  for all  $y \in \text{supp}(Y)$ ,

$$L(\eta_1, \eta_2, \eta_3, g) = \int \left( -S(\mathbf{z}, y) g(y) + \eta_1 g(y) \log\left(\frac{g(y)}{f(y)}\right) + \eta_2 g(y) - \eta_3(y) g(y) \right) dy - \eta_1 \epsilon - \eta_2,$$

where  $\eta_1$  is the Lagrange parameter for the KL constraint,  $\eta_2$  is such that  $g$  integrates to 1, and  $\eta_3(y)$  are such that  $g(y) \geq 0$ , whenever  $f(y) \geq 0$ , and  $g(y) = 0$  otherwise. Using similar steps as in Theorem 1, that is derive the Euler–Lagrange equation, solve it for  $\frac{g(y)}{f(y)}$ , and then impose  $\eta_2, \eta_3$ . Finally setting  $\eta := \frac{1}{\eta_1}$  results in the change of measure

$$\frac{d\mathbb{Q}^\eta}{d\mathbb{P}} := \frac{g(Y)}{f(Y)} = \frac{e^{\eta S(\mathbf{z}, Y)}}{\mathbb{E}[e^{\eta S(\mathbf{z}, Y)}]}.$$

Moreover, it holds that

$$D_{KL}(\mathbb{Q}^\eta \parallel \mathbb{P}) = \eta K'_{S(\mathbf{z}, Y)}(\eta) - \eta K_{S(\mathbf{z}, Y)}(\eta).$$

By similar arguments as in the proof of [Theorem 1](#), the KL constraint is binding and  $\eta^*$  is the unique solution to  $\varepsilon = D_{KL}(\mathbb{Q}^\eta \parallel \mathbb{P})$ . As  $\eta$  and  $\mathbb{Q}^\eta$  both depend on  $z \in A$ , we make this explicit by writing  $\eta(z)$  and  $\mathbb{Q}^{\eta(z)}$  in the statement.  $\square$

[Proposition 3](#) on the properties of the REF also hold in the  $k$ -dimensional setting.

**Proposition 7.** Let  $S : A' \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $A' \subseteq \mathbb{R}^k$ , be a strictly consistent scoring function for a  $k$ -dimensional elicitable functional  $\mathbf{R}$ , then the following holds:

- (i) If  $S(cz, c, y) = c^b S(z, y)$ , where  $cz := (cz_1, \dots, cz_k)$ , for all  $y \in \mathbb{R}$ ,  $c > 0$ , and for some  $b \geq 0$ . Then  $\mathbf{R}$  and  $\bar{\mathbf{R}}^S$  are positive homogeneous of degree 1 for  $A' = (-\infty, 0]^k, [0, \infty)^k$  or  $\mathbb{R}^k$ .
- (ii) If  $A' = \mathbb{R}^k$  and  $S(z - c, y) = S(z, y + c)$  for  $c \in \mathbb{R}$ ,  $c := (c, \dots, c)$  of length  $k$ , then  $\mathbf{R}$  and  $\bar{\mathbf{R}}^S$  are translation invariant.
- (iii) If  $S(y, y) = 0$ , where  $y := (y, \dots, y) \in A'$ , then  $\mathbf{R}(m) = m$  and  $\bar{\mathbf{R}}^S(m) = m$  for all  $m \in \mathbb{R}$ . In particular,  $\mathbf{R}(0) = 0$  and  $\bar{\mathbf{R}}^S(0) = 0$ .

**Proof.** The proof follows using similar arguments as in the proof of [Proposition 3](#).  $\square$

A functional of major interest in the context of risk management is the ES.

**Definition 8 (Expected Shortfall).** The Expected Shortfall, also known as the conditional value-at-risk (CVaR), at tolerance level  $\alpha \in [0, 1]$  for  $X \in \mathcal{L}^\infty$  is defined as

$$\text{ES}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_q(X) dq .$$

Though it is known to not be elicitable alone ([Gneiting, 2011](#)), the pair  $(\text{VaR}_\alpha, \text{ES}_\alpha)$  is 2-elicitable, as shown in [Fissler et al. \(2016\)](#). While many scoring functions exist for this pair, we refer to Theorem C.3 in supplemental of [Nolde and Ziegel \(2017\)](#) for the existence of a  $b$ -homogeneous scoring function for  $(\text{VaR}_\alpha, \text{ES}_\alpha)$ .

**Proposition 8 ( $b$ -homogeneous Scoring Function for  $(\text{VaR}, \text{ES})$ )** [Nolde & Ziegel, 2017](#). For  $\alpha \in (0, 1)$ , the 2-elicitable functional  $(\text{VaR}_\alpha, \text{ES}_\alpha)$  has corresponding scoring functions  $S : \mathbb{R}^3 \rightarrow \mathbb{R}$  of the form

$$S(z_1, z_2, y) = \mathbb{1}_{\{y>z_1\}}(-G_1(z_1) + G_1(y) - G_2(z_2)(z_1 - y)) + (1-\alpha)(G_1(z_1)(z_2 - z_1) + G_2(z_2)), \quad (16)$$

where  $G_1$  is increasing, and  $G_2$  is twice differentiable, strictly increasing, strictly concave and  $G'_2 = G_2$ . If

- (i)  $b \in (0, 1)$ , the only positive homogeneous scoring functions of degree  $b$  and of the form in Eq. (16) are obtained by  $G_1(x) = (d_1 \mathbb{1}\{x \geq 0\} - d_2 \mathbb{1}\{x < 0\})|x|^b - c_0$  and  $G_2(x) = c_1 x^b + c_0$ ,  $x > 0$  where  $c_0 \in \mathbb{R}$ ,  $d_1, d_2 \geq 0$ , and  $c_1 > 0$ .
- (ii)  $b \in (-\infty, 0)$ , the only positive homogeneous scoring function of degree  $b$  and of the form in Eq. (16) are obtained by  $G_1(x) = -c_0$  and  $G_2(x) = -c_1 x^b + c_0$ ,  $x > 0$  where  $c_0 \in \mathbb{R}$  and  $c_1 > 0$ .
- (iii)  $b \in \{0\} \cup (1, +\infty)$ , there are no positively homogeneous scoring functions of the form in Eq. (16) of degree  $b = 0$  or  $b \geq 1$ .

We refer to the supplemental of [Nolde and Ziegel \(2017\)](#) for further details and discussions.

In the next section, we discuss an application of jointly robustifying the  $(\text{VaR}, \text{ES})$  to a reinsurance application.

#### 4.2. Reinsurance application

We consider a reinsurance company who aims to assess the risk associated with its losses. In particular, we are interested in an reinsurance company who has underwritten losses stemming from different

**Table 1**

Distributional assumptions of risk factors of the reinsurance example.

Risk factor	Distribution	Mean	Std
$X_1$	Log-Normal (4.58, 0.19 <sup>2</sup> )	100	20
$X_2$	Log-Normal (4.98, 0.23 <sup>2</sup> )	150	35
$X_3$	Pareto (147.52, 60.65)	150	40

insurance companies. Specifically, the insurance–reinsurance market consists of three insurers and one reinsurer. Each insurance company  $k \in \{1, 2, 3\}$  purchases reinsurance on their business line  $X_k$  with deductible  $d_k$  and limit  $l_k$ . Thus, the total reinsurance loss is

$$Y = \sum_{k=1}^3 \min\{(X_k - d_k)_+, l_k\} .$$

The reinsurer is covering losses between the 60% and the 80% quantile for insurer 1 and 2, and the losses between the 85% and the 95% quantiles of insurer 3, i.e.

$$d_k := F_{X_k}^{-1}(0.6) \quad \text{and} \quad l_k := F_{X_k}^{-1}(0.8), \quad k = 1, 2, \quad \text{and}$$

$$d_3 := F_{X_3}^{-1}(0.85) \quad \text{and} \quad l_3 := F_{X_3}^{-1}(0.95) .$$

The insurance losses  $(X_1, X_2, X_3)$  have marginal distributions described in [Table 1](#) and are dependent through a  $t$ -copula with 4 degrees of freedom and correlation matrix

$$R = \begin{pmatrix} 1 & 0.2 & 0 \\ 0.2 & 1 & 0.8 \\ 0 & 0.8 & 1 \end{pmatrix} .$$

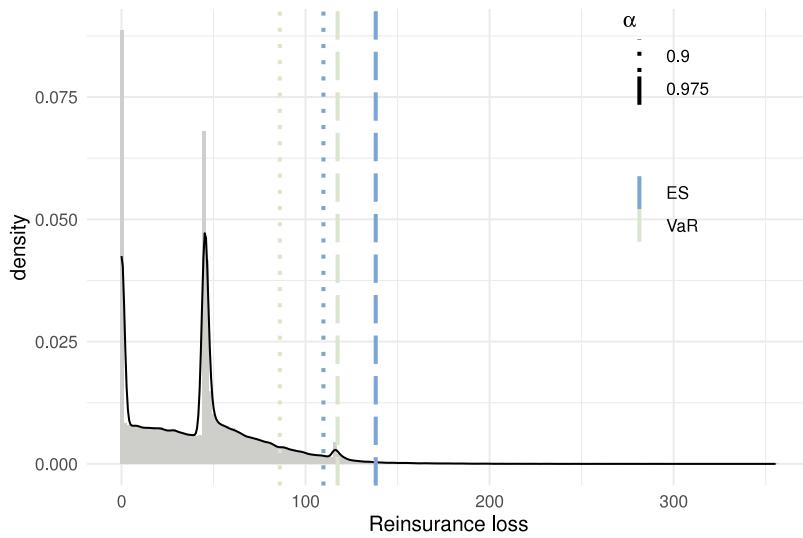
[Fig. 5](#) displays the histogram of the total reinsurance losses  $Y$ , stemming from  $n = 100,000$  samples, and the estimated baseline  $\text{VaR}_\alpha(Y)$  and  $\text{ES}_\alpha(Y)$  for  $\alpha = 0.9, 0.975$ . A smoothed kernel density is displayed as an approximation of the density. In the simulated dataset, the reinsurer's losses are bounded by the sum of the insurers' limit, and the maximal loss for the reinsurer is 448.15. The baseline/empirical VaRs are found to be 86.01 and 109.72 and the baseline/empirical ESs are 117.40 and 138.13, for  $\alpha = 90\%, 97.5\%$  respectively.

Next, we generate a sample of size  $n = 10,000$  of reinsurance losses, and scale the losses by a factor of 0.01. We are able to perform this scaling by using a  $b$ -homogeneous scoring function for the pair  $(\text{VaR}_\alpha, \text{ES}_\alpha)$  and the homogeneity property from [Proposition 7](#).

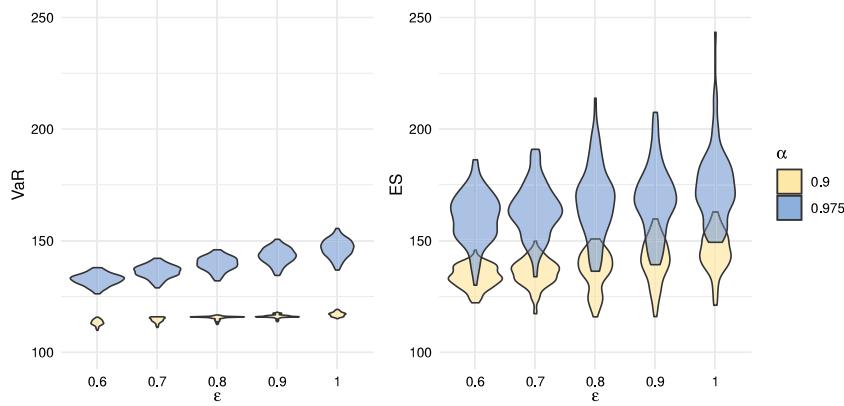
For this sample, we calculate the joint REF  $(\text{VaR}, \text{ES})$ . We do this  $N = 100$  times to illustrate the density of the joint REF VaR and ES. To handle quantile-crossing, we reject loss samples that result in  $\text{VaR}_\alpha > \text{ES}_\alpha$  for any  $\varepsilon = 0.6, 0.7, 0.8, 0.9$ ,  $\alpha = 0.9, 0.975$ . [Fig. 6](#) displays violin diagrams for the joint REF VaR and ES, for the different  $\varepsilon$ 's and  $\alpha$ 's. We find that the joint REF VaR and ES are ordered in  $\varepsilon$ , mirroring the behaviour of the one-dimensional REFs. We also observe that the variance of the REF VaR is significantly smaller than that of the REF ES when found jointly.

#### 5. Application to robust regression

In this section, we extend the REF framework to a regression setting. Distributionally robust regression has been of increasing interest in the machine learning and statistics communities. In the field of statistics, emphasis is placed on extreme value detection and methods that are not too sensitive to extreme values. Significant theory has been developed on this topic, see e.g. [Rousseeuw and Leroy \(2005\)](#) for a comprehensive treatment. These methodologies are now widely applicable to problems in machine learning, though their goals differ. For example, robustness techniques can be used to counter adversarial tests against machine learning algorithms. Specifically for robust regression, the authors ([Shafieezadeh-Abadeh et al., 2015](#)), consider robust logistic regression with uncertainty on both the covariates and the response quantified via a Wasserstein ball, and, in an attempt to



**Fig. 5.** Smoothed density for  $n = 100,000$  simulated reinsurance losses as described in Table 1, with  $\text{VaR}_\alpha$  and  $\text{ES}_\alpha$  for  $\alpha = 90\%, 97.5\%$ .



**Fig. 6.** Densities of robust VaR (left) and robust ES (right) of simulated reinsurance losses for varying  $\epsilon$  and  $\alpha$ . Reference values for VaR are 86.01 and 109.72 and for ES are 117.40 and 138.13, with  $\alpha = 90\%, 97.5\%$  respectively.

mitigate adversarial attacks, Chen and Paschalidis (2018) minimise the worst-case absolute residuals of a discrete regression problem over a Wasserstein ball. Data contamination in DRO problems have likewise been studied, see Xu and Zhang (2021), where authors find sufficient conditions for quantitative robustness of a statistical estimator of DRO models.

The motivation for robustness in risk management differs from these communities. Of concern are extreme events, where data may be sparse due to the rarity of events, which a risk measure should ideally be capturing.

### 5.1. Robust regression coefficients

Here we propose an approach where we consider distributional uncertainty jointly in both the covariate and the response variable, and where the uncertainty is characterised via the KL-divergence. For this let  $\mathbf{X} := (X_1, \dots, X_m)$  be the  $m$ -dimensional covariates such that each component is in  $\mathcal{L}^\infty$ , i.e.,  $X_k \in \mathcal{L}^\infty$ , for all  $k = 1 \dots, m$ , and let  $Y \in \mathcal{L}^\infty$  be a univariate response.

For an elicitable functional  $R$  with strictly consistent scoring function  $S$ , we make the classical regression assumption that

$$R(Y|\mathbf{X} = \mathbf{x}) = \beta_1 x_1 + \dots + \beta_m x_m,$$

where  $R(Y|\mathbf{X} = \mathbf{x})$  denotes the functional  $R$  evaluated on the conditional cdf of  $Y$  given  $\mathbf{X} = \mathbf{x}$ . The parameters  $\beta := (\beta_1, \dots, \beta_m)$ , the

regression coefficients, are estimated via solving the sample-version of the following minimisation problem

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^m}{\operatorname{argmin}} \mathbb{E}[S(\beta^\top \mathbf{X}, Y)]. \quad (17)$$

As we allow  $\beta \in \mathbb{R}^m$ , in this section, we only consider scoring functions defined on  $\mathbb{R}^2$ , i.e.  $S : \mathbb{R}^2 \rightarrow [0, \infty)$ . Moreover, for simplicity, we assume that the functional  $R(Y|\mathbf{X} = \mathbf{x})$  is linear in the covariates, though the results can be adapted to include link functions. For the choice of the mean functional  $R$  and the squared loss  $S(z, y) = (y - z)^2$ , we recover the usual linear regression. The classical quantile regression follows by setting the scoring function to be the pinball loss, i.e.,  $S(z, y) = (\mathbb{1}_{\{y \leq z\}} - \alpha)(z - y)$ , which is strictly consistent for the  $\alpha$ -quantile,  $\alpha \in (0, 1)$ . Similarly one can obtain expectile and ES regression.

We propose to robustify the regression coefficients  $\hat{\beta}$  of Eq. (17) by accounting for uncertainty in the joint distribution between the response and the covariates, that is of  $(\mathbf{X}, Y)$ .

**Definition 9 (Robust Regression Coefficients).** Let  $R$  be an elicitable functional with strictly consistent scoring function  $S : \mathbb{R}^2 \rightarrow [0, \infty)$  and  $\epsilon \geq 0$ . Then the **robust regression coefficients** are given by

$$\bar{\beta}^S := \underset{\beta \in \mathbb{R}^m}{\operatorname{argmin}} \sup_{Q \in Q_\epsilon} \mathbb{E}^Q[S(\beta^\top \mathbf{X}, Y)], \quad (18)$$

where the uncertainty set  $Q_\epsilon$  is given in Eq. (7).

The representation of the inner optimisation problem in (18) also holds in the robust regression setting under a similar assumption on the tolerance distance  $\varepsilon$ .

**Assumption 3 (Maximal Kullback–Leibler Distance).** Define  $p(\beta) := \mathbb{P}(S(\beta^\top \mathbf{X}, Y) = \text{ess sup } S(\beta^\top \mathbf{X}, Y))$ . Then one of the following holds:

- (i) If  $p(\beta) = 0$  for all  $\beta \in \mathbb{R}^m$ , then the tolerance distance satisfies  $\varepsilon \in [0, \infty)$ .
- (ii) If  $p(\beta) > 0$  for some  $\beta \in \mathbb{R}^m$ , then the tolerance distance  $\varepsilon$  satisfies

$$0 \leq \varepsilon < \log\left(\frac{1}{p(\beta)}\right) \quad \text{for all } \beta \in \mathbb{R}^m.$$

**Corollary 2.** Let  $S$  be a strictly consistent scoring function for  $R$ , and  $\varepsilon$  such that **Assumption 3** is satisfied. Then, for  $m$ -dimensional covariates  $\mathbf{X}$ , satisfying  $X_k \in \mathcal{L}^\infty$ ,  $k = 1, \dots, m$ , and a univariate response  $Y \in \mathcal{L}^\infty$ , the robust regression coefficients have representation

$$\bar{\beta}^S = \underset{\beta \in \mathbb{R}^m}{\operatorname{argmin}} \frac{\mathbb{E}[S(\beta^\top \mathbf{X}, Y)e^{\eta^*(\beta)S(\beta^\top \mathbf{X}, Y)}]}{\mathbb{E}[e^{\eta^*(\beta)S(\beta^\top \mathbf{X}, Y)}]},$$

where for each  $\beta \in \mathbb{R}^m$ ,  $\eta^*(\beta) \geq 0$  is the unique solution to  $\varepsilon = D_{KL}(\mathbb{Q}^{\eta(\beta)} \parallel \mathbb{P})$  with

$$\frac{d\mathbb{Q}^{\eta(\beta)}}{d\mathbb{P}} := \frac{e^{\eta(\beta)S(\beta^\top \mathbf{X}, Y)}}{\mathbb{E}[e^{\eta(\beta)S(\beta^\top \mathbf{X}, Y)}]}.$$

**Proof.** The proof is similar to that of **Theorem 1**. For simplicity, we assume that  $(\mathbf{X}, Y)$  has joint pdf  $f$  under the baseline measure  $\mathbb{P}$ . For simplicity we let  $(\mathbf{X}, Y)$  be continuously distributed with density  $f(\cdot, \cdot)$  under  $\mathbb{P}$ . Analogous to **Theorem 1**, this can be generalised to  $(\mathbf{X}, Y)$  being discrete or continuously distributed except for a countable number of points.

Then, the inner optimisation problem of (18) can be written as an optimisation problem over the joint density of  $(\mathbf{X}, Y)$  as follows

$$\sup_{g: \mathbb{R}^{m+1} \rightarrow \mathbb{R}} \iint S(\beta^\top \mathbf{x}, y)g(\mathbf{x}, y)dy d\mathbf{x}, \quad \text{subject to}$$

$$\begin{aligned} & \iint \frac{g(\mathbf{x}, y)}{f(\mathbf{x}, y)} \log\left(\frac{g(\mathbf{x}, y)}{f(\mathbf{x}, y)}\right) \\ & \times f(\mathbf{x}, y) dy d\mathbf{x} \leq \varepsilon, \\ & \iint g(\mathbf{x}, y) dy d\mathbf{x} = 1, \quad \text{and} \\ & g(\mathbf{x}, y) \geq 0, \quad \text{for all } \mathbf{x} \in \mathbb{R}^m, \\ & y \in \text{supp}(Y). \end{aligned}$$

This optimisation problem admits the Lagrangian, with Lagrange parameters  $\eta_1, \eta_2 \geq 0$  and  $\eta_3(\mathbf{x}, y) \geq 0$  for all  $\mathbf{x} \in \mathbb{R}^m$ ,  $y \in \text{supp}(Y)$ ,

$$\begin{aligned} L(\eta_1, \eta_2, \eta_3, g) = & \iint \left( -S(\beta^\top \mathbf{x}, y)g(\mathbf{x}, y) + \eta_1 g(\mathbf{x}, y) \log\left(\frac{g(\mathbf{x}, y)}{f(\mathbf{x}, y)}\right) \right. \\ & \left. + \eta_2 g(\mathbf{x}, y) - \eta_3(\mathbf{x}, y)g(\mathbf{x}, y) \right) dy d\mathbf{x} \\ & - \eta_1 \varepsilon - \eta_2, \end{aligned}$$

where  $\eta_1$  is the Lagrange parameter for the KL constraint,  $\eta_2$  is such that  $g$  integrates to 1, and  $\eta_3(\mathbf{x}, y)$  are such that  $g(\mathbf{x}, y) \geq 0$ , whenever  $f(\mathbf{x}, y) \geq 0$ , and  $g(\mathbf{x}, y) = 0$  otherwise. Using similar steps as in **Theorem 1**, that is derive the Euler–Lagrange equation, solve it for  $\frac{g(\mathbf{x}, y)}{f(\mathbf{x}, y)}$ , and then impose  $\eta_2, \eta_3$ . Finally setting  $\eta := \frac{1}{\eta_1}$  results in the change of measure

$$\frac{d\mathbb{Q}^\eta}{d\mathbb{P}} := \frac{g(\mathbf{X}, Y)}{f(\mathbf{X}, Y)} = \frac{e^{\eta S(\beta^\top \mathbf{X}, Y)}}{\mathbb{E}[e^{\eta S(\beta^\top \mathbf{X}, Y)}]}.$$

Moreover, it holds that

$$D_{KL}(\mathbb{Q}^\eta \parallel \mathbb{P}) = \eta K'_{S(\beta^\top \mathbf{X}, Y)}(\eta) - \eta K_{S(\beta^\top \mathbf{X}, Y)}(\eta).$$

**Table 2**

Results of robust linear regression for models A, B, and C. The parameter  $\beta_0$  corresponds to the intercept,  $\beta_1$  is the slope of the regression line, and MSE the mean squared error.

$\varepsilon$	Model A			Model B			Model C		
	$\beta_0$	$\beta_1$	MSE	$\beta_0$	$\beta_1$	MSE	$\beta_0$	$\beta_1$	MSE
0	0.05	0.88	0.016	0.18	0.69	0.036	0.24	0.48	0.054
1	0.18	0.68	0.020	0.40	0.33	0.044	0.35	0.21	0.062
5	0.26	0.49	0.028	0.37	0.28	0.051	0.38	0.14	0.065
10	0.29	0.42	0.030	0.37	0.25	0.051	0.32	0.19	0.079

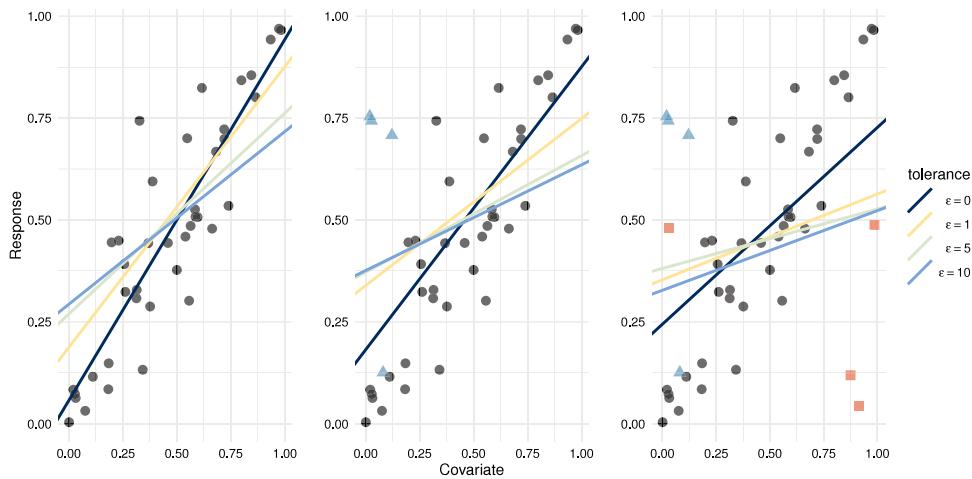
By the similar arguments as in the proof of **Theorem 1**, the KL constraint is binding and  $\eta^*$  is the unique solution to  $\varepsilon = D_{KL}(\mathbb{Q}^\eta \parallel \mathbb{P})$ . As  $\eta$  and  $\mathbb{Q}^\eta$  both depend on  $\beta \in \mathbb{R}^m$ , we make this explicit by writing  $\eta(\beta)$  and  $\mathbb{Q}^{\eta(\beta)}$  in the statement.  $\square$

We consider a numerical case study where we compare three different models for the joint dependence structure of a bivariate covariate and response pair  $(X, Y)$ , as detailed below. Here the dimension of the covariates is 1, i.e.,  $m = 1$ . Motivated by a dataset being contaminated with confounding or extreme values, we prescribe  $(X, Y)$  as having a Gumbel(5) copula with uniform marginals, as the “reference” samples without any confounding values (model A). The choice of Gumbel(5) copula yields a Kendall’s tau of 0.8, giving a suitable linear relationship to test linear regression. The “contaminated” models (models B and C) have additional to the data points of model A, samples of  $(\tilde{X}, \tilde{Y})$ , that we identify as confounding values, where  $(\tilde{X}, \tilde{Y})$  have uniform marginal distributions and an independent copula.

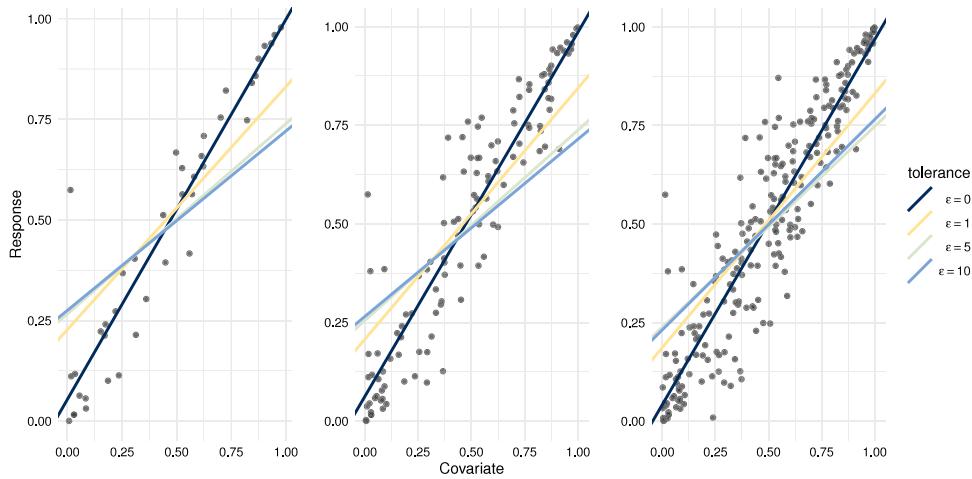
**Example 4 (Data Contamination).** In particular, the dataset we consider is constructed in the following way: we sample from the Gumbel(5) copula to form the uncontaminated dataset in model A, then augment model A with 4 independent extreme values to result in model B, and then again augment model B with 4 additional extreme values to generate model C. The model B consists of 9% independent extreme values, and model C of 18%. We fit a robust linear regression, with varying tolerances, to these 3 models, using the squared loss scoring function  $S(z, y) = (z - y)^2$ . We further calculate the traditional linear regression, which coincides with tolerance  $\varepsilon = 0$ . **Fig. 7** displays the datasets for models A, B, and C in the panels from left to right, and the robust regression lines for varying tolerances  $\varepsilon$ . In the leftmost panel, we have the uncontaminated dataset model A, where we find that the robust linear regression has steeper regression lines for smaller tolerances. The centre panel displays the data and linear regression lines from this centre panel behave similarly to those of model A, in that steeper regression lines correspond to smaller  $\varepsilon$ , though the confounding of the sample reduces the overall steepness of the regression lines. In the left and middle panels, the slope of the robust regression decreases as the uncertainty tolerance  $\varepsilon$  increases, which is different from the right panel where we do not observe a clear ordering. The right panel displays the robust linear regression of the most contaminated dataset, model C. In this scenario, the robust linear regression is significantly flatter for all values of  $\varepsilon$  compared to the linear regression lines with  $\varepsilon = 0$ . Clearly, including extreme values reduces the slope of the regression line. Moreover, we observe that the larger the tolerance distance  $\varepsilon$ , meaning as we allow for more uncertainty, the flatter the regression lines.

For completeness, we report in **Table 2** the regression coefficients and mean squared errors (MSEs) of the robust linear regressions of models A, B, and C. As observed in **Fig. 7**, the larger  $\varepsilon$  the smaller the slope of the regression line, i.e.  $\beta_1$ .

**Example 5 (Sensitivity to Sample Size).** Another situation in which one may want to robustify linear regression is when the data sample is sparse, e.g. to introduce uncertainty to prevent model overfit. Here, we



**Fig. 7.** Fitted robust linear regressions lines for models A, B, C (left to right) and for different  $\epsilon = 0, 1, 5, 10$ . The blue pyramid and the red squares correspond to the extreme values in models B and C. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 8.** Fitted robust linear regression lines for model A with sample size  $n = 40, 80, 120$  (left to right) and for with different  $\epsilon = 0, 1, 5, 10$ .

consider the behaviour of the robust regression for increasing sample sizes. We generate covariate and response pairs having Gumbel(5) copula and uniform marginals, i.e. Model A. The first dataset consists of 40 samples, then we increase the dataset to 80 samples by adding another 40 samples, resulting in the second dataset. Finally, for the third dataset, we add another 40 samples to obtain a sample size of 120. **Fig. 8** displays model A for sample sizes  $n = 40, 80, 120$ , from left to the right panel. Again, we observe that the smaller  $\epsilon$ , the steeper the slope of the regression lines.

**Table 3** reports the regression coefficients and MSEs of the robust linear regressions for the  $n = 40, 80, 120$  sample size datasets. As expected, for linear regression with  $\epsilon = 0$ , we have the smallest MSE, as it, by definition, is the solution to minimising the MSE. The MSE of the regular linear regression remains approximately the same for the three sample sizes, while those of the robust linear regression have a larger deviation. Moreover, for each value of  $\epsilon$ , the MSE decreases with increasing sample size.

## 5.2. Spanish motor vehicle insurance case study

We perform expectile regression on a real automotive insurance dataset published in Segura-Gisbert et al. (2024). We consider the response to be the premium paid, and the covariates to be the value of the vehicle and the claims made, in tens of thousands of Euro and hundreds

**Table 3**

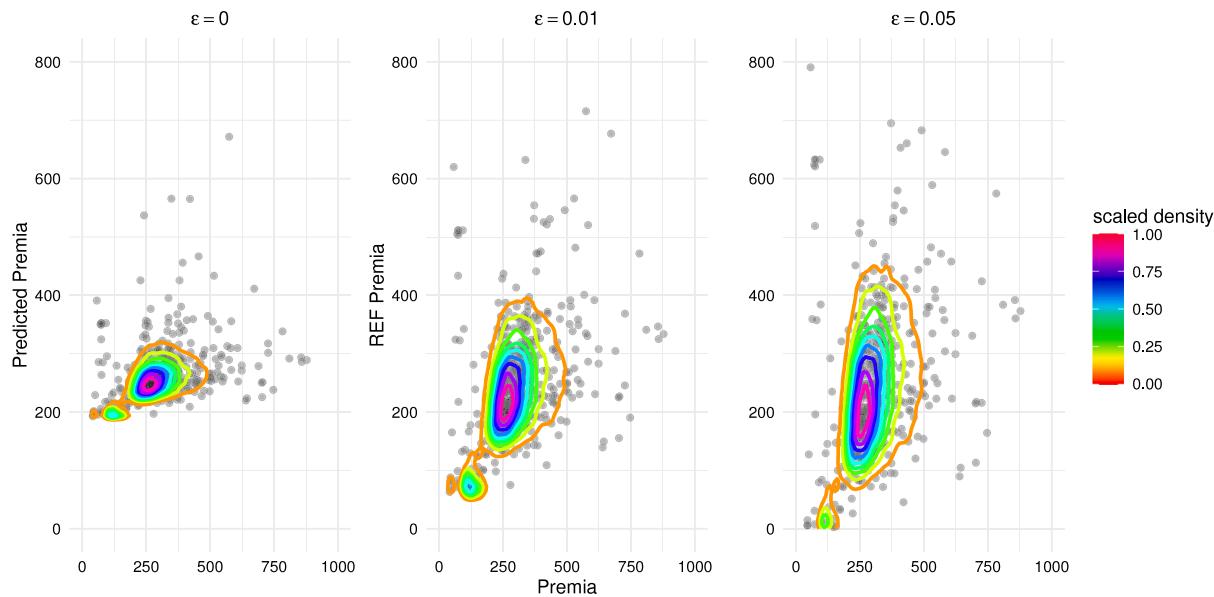
Results of robust linear regression for different dataset sizes  $n = 40, 80$ , and  $120$ . The parameter  $\beta_0$  corresponds to the intercept,  $\beta_1$  is the slope of the regression line, and MSE the mean squared error.

$\epsilon$	$n = 40$			$n = 80$			$n = 120$		
	$\beta_0$	$\beta_1$	MSE	$\beta_0$	$\beta_1$	MSE	$\beta_0$	$\beta_1$	MSE
0	0.05	0.94	0.011	0.06	0.92	0.012	0.41	0.92	0.01
1	0.22	0.60	0.023	0.20	0.63	0.020	0.18	0.64	0.019
5	0.26	0.47	0.033	0.25	0.48	0.031	0.23	0.50	0.027
10	0.27	0.44	0.033	0.36	0.44	0.031	0.23	0.53	0.027

of thousand Euro, respectively. We bootstrap a sample of size  $n = 1000$  30 times, and find the REF expectile regression coefficients  $\beta_0, \beta_1, \beta_2$  and baseline expectile regression coefficients, at level  $\tau = 0.8$ . For the REF expectile regression coefficients, we use the  $b$ -homogeneous scoring function for the expectile, as given in [Proposition 6](#), with  $b = 2$ . We compare this to expectile regression as implemented in the R package, `expectreg` (Otto-Sobotka et al., 2024), which corresponds to  $\epsilon = 0$ .

In [Table 4](#), we report the regression coefficients and MSE for traditional expectile regression ( $\epsilon = 0$ ), and using the REF methodology with  $\epsilon = 0.01, 0.05$ . We find that using the REF in this circumstance reduced the MSE. This is in contrast to [Example 4](#), where canonical linear regression minimises the MSE by construction.

To create [Fig. 9](#), we chose 500 data points are random from the full dataset. The predicted and REF premia are then calculated using the



**Fig. 9.** Canonical expectile regression predicted premia and REF predicted premia against actualised premia with  $\epsilon = 0, 0.01, 0.05$  from left to right. Grey points correspond to how real data points are mapped to predicted premia.

**Table 4**

Results of bootstrapped robust expectile regression for automotive insurance data. The parameter  $\beta_0$  corresponds to the intercept,  $\beta_1$  to the claims,  $\beta_2$  to the value of the vehicle, and MSE the mean squared error.

$\epsilon$	$\beta_0$	$\beta_1$	$\beta_2$	MSE
0	0.0187	0.0840	0.0383	0.00032
0.01	0.0049	0.0455	0.1076	0.00025
0.05	-0.0041	0.0334	0.1570	0.00029

actualised covariates of this sample and plotted against the true premia. Again we interpret increasing  $\epsilon$  as an increase in the joint uncertainty between all covariates and the response. This is supported in Fig. 9, where under no uncertainty (left panel), the predicted premia clearly exhibit a dense cluster with a high peak, indicating that the predicted premia are highly concentrated near its median of 259. When using the REF, as in the centre and right panels, we see that, while there is a concentration, the density flattens as  $\epsilon$  increases, indicating that there is more variability in the REF premia compared to when  $\epsilon = 0$ .

## 6. Conclusion

This paper proposes a new robustification of elicitable functionals, the REF, by incorporating uncertainty prescribed by the KL divergence. Mathematically, the REF is the argmin of an extremal score, where the extremal score is the largest expected score over a set of alternative probability measures that lie within a KL ball. Thus the REF takes the form of a minimax optimisation problem. This methodology is distinct from classical worst-case risk measurement, wherein the risk functional is given as the supremum over all alternative probability measures within the KL ball, of the minimisers of the expected scores. For the REF, we show that the constraint on the uncertainty region is binding, and characterise conditions for the existence and uniqueness.

Since the REF depends on the choice of scoring function, we explore this choice by using  $b$ -homogeneous scoring functions, show that these families of scoring functions preserve desirable properties for the REF, and illustrate them using Murphy diagrams.

We extend the REF and its representation results to two settings: to  $k$ -dimensional elicitable functionals, and to functional regression application. In the  $k$ -dimensional setting, we consider an application to reinsurance by demonstrating the behaviour of the joint (VaR, ES)

REF to a synthetic reinsurance dataset. In the robust regression setting, we explore the behaviour of the REF under three scenarios: a data contamination problem, a sample size scenario, and a data driven automotive insurance example.

## CRediT authorship contribution statement

**Kathleen E. Miao:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Conceptualization. **Silvana M. Pesenti:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Funding acquisition, Conceptualization.

## Acknowledgements

The authors thank Fabio Gomez for stimulating discussions. SP gratefully acknowledges support from the Natural Sciences and Engineering Research Council of Canada, Canada (grants DGECR-2020-00333 and RGPIN-2020-04289) and from the Canadian Statistical Sciences Institute (CANSSI).

## References

- Albrecher, H., Beirlant, J., & Teugels, J. (2017). *Wiley series in probability and statistics, Reinsurance: Actuarial and statistical aspects*. Wiley, ISBN: 9780470772683.
- Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203–228. <http://dx.doi.org/10.1111/1467-9965.00068>.
- Bellini, F., & Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance*, 15(5), 725–733. <http://dx.doi.org/10.1080/14697688.2014.946555>.
- Bellini, F., Klar, B., Müller, A., & Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics & Economics*, [ISSN: 0167-6687] 54, 41–48. <http://dx.doi.org/10.1016/j.inmatheco.2013.10.015>.
- Bernard, C., Pesenti, S. M., & Vanduffel, S. (2024). Robust distortion risk measures. *Mathematical Finance*, 34(3), 774–818. <http://dx.doi.org/10.1111/mafi.12414>.
- Blanchet, J., Lam, H., Tang, Q., & Yuan, Z. (2019). Robust actuarial risk analysis. *North American Actuarial Journal*, 23(1), 33–63. <http://dx.doi.org/10.1080/10920277.2018.1504686>.
- Cai, J., Liu, F., & Yin, M. (2024). Worst-case risk measures of stop-loss and limited loss random variables under distribution uncertainty with applications to robust reinsurance. *European Journal of Operational Research*, [ISSN: 0377-2217] 318(1), 310–326. <http://dx.doi.org/10.1016/j.ejor.2024.03.016>.
- Calafiore, G. C. (2007). Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3), 853–877. <http://dx.doi.org/10.1137/060654803>.

- Chen, R., & Paschalidis, I. C. (2018). A robust learning approach for regression models based on distributionally robust optimization. *Journal of Machine Learning Research*, 19(13), 1–48, URL <http://jmlr.org/papers/v19/17-295.html>.
- Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, 93–125, URL <https://www.jstor.org/stable/24303995>.
- Ehm, W., Gneiting, T., Jordan, A., & Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 78(3), 505–562, URL <10.1111/rssb.12154>.
- Embrechts, P., Mao, T., Wang, Q., & Wang, R. (2021). Bayes risk, elicibility, and the expected shortfall. *Mathematical Finance*, 31(4), 1190–1217. <http://dx.doi.org/10.1111/mafi.12313>.
- Embrechts, P., Puccetti, G., & Rüschedorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking & Finance*, [ISSN: 0378-4266] 37(8), 2750–2764. <http://dx.doi.org/10.1016/j.jbankfin.2013.03.014>.
- Embrechts, P., Wang, B., & Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, 19(4), 763–790. <http://dx.doi.org/10.1007/s00780-015-0273-z>, ISSN 0949-2984, 1432-1122.
- Fissler, T., & Pesenti, S. M. (2023). Sensitivity measures based on scoring functions. *European Journal of Operational Research*, [ISSN: 0377-2217] 307(3), 1408–1423. <http://dx.doi.org/10.1016/j.ejor.2022.10.002>.
- Fissler, T., & Ziegel, J. F. (2016). Higher order elicibility and Osband's principle. *Annals of Statistics*, 44(4), 1680–1707. <http://dx.doi.org/10.1214/16-AOS1439>.
- Fissler, T., & Ziegel, J. F. (2021). On the elicibility of range value at risk. *Statistics & Risk Modeling*, 38(1–2), 25–46. <http://dx.doi.org/10.1515/strm-2020-0037>.
- Fissler, T., Ziegel, J. F., & Gneiting, T. (2016). Expected shortfall is jointly elicitable with Value-at-Risk: implications for backtesting. *Risk Magazine*, 58–61. <http://dx.doi.org/10.48550/arXiv.1507.00244>.
- Ghaoui, L. E., Oks, M., & Oustry, F. (2003). Worst-case Value-At-Risk and robust portfolio optimization: A Conic Programming Approach. *Operations Research*, 51(4), 543–556. <http://dx.doi.org/10.1287/opre.51.4.543.16101>.
- Glasserman, P., & Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance*, 14(1), 29–58. <http://dx.doi.org/10.1080/14697688.2013.822989>.
- Gneiting, T. (2011). Making and Evaluating Point Forecasts. *Journal of the American Statistical Association*, 106(494), 746–762, URL <https://www.jstor.org/stable/41416407>.
- He, X. D., Kou, S., & Peng, X. (2022). Risk measures: Robustness, elicibility, and backtesting. *Annual Review of Statistics and its Application*, [ISSN: 2326-831X] 9(Volume 9, 2022), 141–166. <http://dx.doi.org/10.1146/annurev-statistics-030718-105122>.
- Hu, Z., & Hong, L. J. (2013). Kullback-Leibler divergence constrained distributionally robust optimization. Available at Optimization Online, 1(2), 9.
- Kou, S., & Peng, X. (2016). On the measurement of economic tail risk. *Operations Research*, 64(5), 1056–1072, URL <http://www.jstor.org/stable/26153487>, ISSN 0030364X, 15265463.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86.
- Lassance, N., & Vrins, F. (2023). Portfolio selection: A target-distribution approach. *European Journal of Operational Research*, [ISSN: 0377-2217] 310(1), 302–314. <http://dx.doi.org/10.1016/j.ejor.2023.02.014>.
- Milgrom, P., & Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2), 583–601. <http://dx.doi.org/10.1111/1468-0262.00296>.
- Nolde, N., & Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *Annals of Applied Statistics*, 11(4), 1833–1874. <http://dx.doi.org/10.1214/17-AOAS1041>.
- Otto-Sobotka, F., Spiegel, E., Schnabel, S., Schulze Waltrup, L., Eilers, P., Kauermann, G., & Kneib, T. (2024). expectreg: Expectile and quantile regression. URL <https://CRAN.R-project.org/package=expectreg>, R package version 0.53.
- Patton, A. J. (2011). Data-based ranking of realised volatility estimators. *Journal of Econometrics*, 161(2), 284–303. <http://dx.doi.org/10.1016/j.jeconom.2010.12.010>.
- Pesenti, S. M., Millosovich, P., & Tsanakas, A. (2016). Robustness regions for measures of risk aggregation. *Dependence Modeling*, 4(1), Article 000010151520160020. <http://dx.doi.org/10.1515/demo-2016-0020>.
- Pesenti, S. M., Millosovich, P., & Tsanakas, A. (2019). Reverse sensitivity testing: What does it take to break the model? *European Journal of Operational Research*, 274(2), 654–670, URL <10.1016/j.ejor.2018.10.003>.
- Rousseuw, P., & Leroy, A. (2005). *Wiley series in probability and statistics, Robust regression and outlier detection*. Wiley, ISBN: 9780471725374.
- Rüschedorf, L., Vanduffel, S., & Bernard, C. (2024). *Model risk management: Risk bounds under uncertainty*. Cambridge University Press, ISBN: 9781009367189, <http://dx.doi.org/10.1017/9781009367189>.
- Segura-Gisbert, J., Lledó, J., & Pavía, J. M. (2024). Dataset of an actual motor vehicle insurance portfolio. *European Actuarial Journal*, 1–13.
- Shafieezadeh-Abadeh, S., Esfahani, P. M., & Kuhn, D. (2015). Distributionally robust logistic regression. In *Proceedings of the 28th international conference on neural information processing systems - volume 1 NIPS '15*, (pp. 1576–1584). Cambridge, MA, USA: MIT Press, <http://dx.doi.org/10.48550/arXiv.1509.09259>.
- Steinwart, I., Pasin, C., Williamson, R. C., & Zhang, S. (2014). Elicitation and identification of properties. In *Annual conference computational learning theory*. URL <https://api.semanticscholar.org/CorpusID:15711300>.
- van Erven, T., & Harremoës, P. (2012). Rényi divergence and Kullback-Leibler divergence.
- Wang, R., & Ziegel, J. F. (2015). Elicitable distortion risk measures: A concise proof. *Statistics & Probability Letters*, 100, 172–175. <http://dx.doi.org/10.1016/j.spl.2015.02.004>.
- Xu, H., & Zhang, S. (2021). Quantitative statistical robustness in distributionally robust optimization models. *Pacific Journal of Optimization Special Issue*, URL <https://optimization-online.org/wp-content/uploads/2021/12/8709.pdf>.
- Ziegel, J. F. (2016). Coherence and elicibility. *Mathematical Finance*, 26(4), 901–918. <http://dx.doi.org/10.1111/mafi.12080>.