

# Robustness in the Optimization of Risk Measures

Paul Embrechts\*      Alexander Schied†      Ruodu Wang‡

March 2, 2021

## Abstract

We study issues of robustness in the context of Quantitative Risk Management and Optimization. We develop a general methodology for determining whether a given risk measurement related optimization problem is robust, which we call “robustness against optimization”. The new notion is studied for various classes of risk measures and expected utility and loss functions. Motivated by practical issues from financial regulation, special attention is given to the two most widely used risk measures in the industry, Value-at-Risk (VaR) and Expected Shortfall (ES). We establish that for a class of general optimization problems, VaR leads to non-robust optimizers whereas convex risk measures generally lead to robust ones. Our results offer extra insight on the ongoing discussion about the comparative advantages of VaR and ES in banking and insurance regulation. Our notion of robustness is conceptually different from the field of robust optimization, to which some interesting links are derived.

**Keywords:** robustness, Value-at-Risk, Expected Shortfall, optimization, financial regulation

---

\*RiskLab, Department of Mathematics, and ETH Risk Center, ETH Zurich, 8092 Zurich, Switzerland. Email: [embrechts@math.ethz.ch](mailto:embrechts@math.ethz.ch)

†Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada. Email: [aschied@uwaterloo.ca](mailto:aschied@uwaterloo.ca)

‡Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada. Email: [wang@uwaterloo.ca](mailto:wang@uwaterloo.ca).

When a measure becomes a target, it ceases to be a good measure.

– *Goodhart’s law, paraphrased by Strathern (1997)*

## 1 Introduction

The main focus of this paper is the study of robustness properties of optimization procedures within Quantitative Risk Management (QRM). For this, we introduce a novel general framework, which at the same time is conceptually intuitive and mathematically challenging. A key and, as we will highlight in the paper, novel question concerns the influence of the choice of the underlying objective on the resulting robustness properties in risk optimization. In particular, we are interested in the two most popular regulatory risk measures, the Value-at-Risk (VaR) and the Expected Shortfall (ES), and their robustness properties in the context of risk optimization.

In QRM, the concept of robustness for risk measures is traditionally studied at the level of objective functionals without involving optimization problems; see [Cont et al. \(2010\)](#), [Kou et al. \(2013\)](#), [Krättschmer et al. \(2014, 2017\)](#), [Embrechts et al. \(2015\)](#), and the references therein. In the literature on robust optimization (see e.g. [Ben-Tal et al. \(2009\)](#)), model uncertainty is typically incorporated through modifying the objective functional or the constraints.

The paper [Cont et al. \(2010\)](#) compares the qualitative robustness of VaR and coherent risk measures; the authors conclude that VaR is better in their context. Some later papers, e.g. [Embrechts et al. \(2015\)](#) and [Krättschmer et al. \(2014, 2017\)](#), put the corresponding arguments into a different perspective, showing that ES also has certain desirable robustness properties. Both streams of research assumed that both VaR and, say, ES are applied to the same financial position. In reality, however, the regulatory choice of a particular risk measure creates certain incentives, just like any other aspect of regulation. These incentives become effective even before that risk measure has ever been applied in a risk management context. For instance, once a specific risk measure has been chosen, portfolios will be optimized with respect to that risk measure. Thus, in reality, VaR and ES will typically not be applied to the same position, and so one cannot decouple the technical properties of a chosen risk measure from the incentives it creates. In other words, a risk measure as a standalone function may be robust, but fail to have desirable robustness properties when this measure is used within an optimization context.

In our paper, we make a first attempt of taking the incentives created by the choice of a risk measure into account when assessing the risk measure’s robustness properties. In doing so, we arrive at completely different, and perhaps somewhat surprising, conclusions concerning robustness properties than the previous literature. To briefly illustrate our ideas, suppose that a risk factor is represented by a random variable  $X$  arising from a stochastic model (later in the

paper,  $X$  will denote a random vector). An investor has to optimize her position according to the best of her knowledge, and hence we shall refer to  $X$  as the best-of-knowledge model, and the true model, denoted by  $Z$ , is typically unknowable. Ideally, a good model  $X$  is statistically close to  $Z$  in a sense to be made clear later. Based on the best-of-knowledge model  $X$  and an objective functional  $\rho$ , an optimized position is chosen as a function  $g(X)$  of  $X$ . Whereas the position  $g(X)$  may have a desirable objective value  $\rho(g(X))$ , this does not guarantee that  $\rho(g(Z))$  is also desirable if  $Z$  is “slightly” different from  $X$ . In the absence of a perfect model, which almost always is the case in financial applications, this issue becomes crucially important.

Our motivation can be informally illustrated by Figure 1, which describes a situation in which the investor faces a Pareto-distributed risk with unknown parameter  $\theta$ .<sup>1</sup> The investor takes a model  $X$  with an estimated parameter  $\hat{\theta}$  and optimizes  $\rho(g(X))$  over a certain class of functions  $g$ ; details are explained in the caption of Figure 1. We compute the actual risk  $\rho(g_X(Z))$  faced by the investor for  $\rho$  chosen as VaR at level 0.99 and ES at level 0.975 as in BCBS (2016), where  $g_X$  denote the corresponding optimizing functions (unique in the case of VaR) using the model  $X$  and  $\rho$ . As one can see from Figure 1, the *perceived* risk value  $\rho(g_X(X))$  (see Section 2.3 for its interpretation) is similar for  $\rho$  being VaR and ES; this also holds for the *actual* risk value  $\rho(g_X(Z))$  as long as  $\theta > \hat{\theta}$ , meaning that  $X$  is an overestimate of the true loss  $Z$ . If, however, the true loss is slightly underestimated, the actual VaR of  $g_X(Z)$  is substantially higher than  $\rho(g_X(X))$ , whereas the actual ES remains almost flat. The intuitive explanation of this phenomenon is that under VaR it is optimal to concentrate all tail risk on an event whose probability is so small that it does not affect the measured risk. If, however, the true probability of that event is underestimated by the model, then the tail risk suddenly does become significant. Simulation results show that, if  $\hat{\theta}$  is computed from a maximum likelihood estimator based on iid sample points for the a true parameter, then the mean square error measured by the difference between  $\rho(g_X(X))$  and  $\rho(g_X(Z))$  is large for the case of VaR, and it is tiny for the case of ES (see Section 7); this sharp contrast is expected from Figure 1.

In the present paper, we put the above observation into a rigorous quantitative framework for general risk measures and optimization problems. The contribution and the structure of the paper are outlined below. In Section 2, we introduce the theoretical framework of robustness properties of risk measures in the context of optimization problems, referred to as “robustness against optimization” throughout the paper.<sup>2</sup> The framework is quite general and it includes many practical problems in various fields of applications, not necessarily confined to finance and insurance. Keeping the problem of risk measures in mind, a class of functional optimization

---

<sup>1</sup>The Pareto( $\theta$ ) distribution function  $F$  with parameter  $\theta > 0$  is specified as  $F(x) = 1 - x^{-\theta}$ ,  $x \geq 1$ .

<sup>2</sup>We thank the Area Editor, Paul Glasserman for suggesting this terminology.

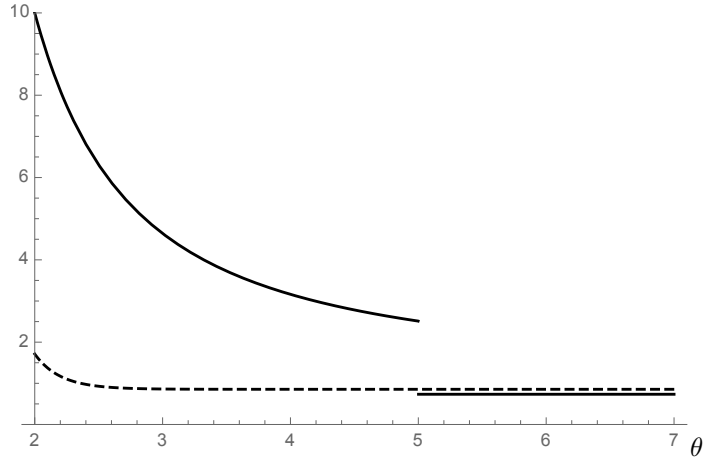


Figure 1: This plot shows  $\rho(g_X(Z))$  for  $\rho = \text{VaR}_{0.99}$  (solid) and  $\rho = \text{ES}_{0.975}$  (dashed), if  $Z$  is Pareto distributed with parameter  $\theta$ . We assume that  $X$  has a Pareto distribution with parameter  $\hat{\theta} = 5$  and  $g_X$  minimizes  $\rho(g(X))$  within the class of all measurable functions  $g$  satisfying the inequality  $0 \leq g(x) \leq x$  for all  $x \geq 0$  and the budget constraint  $\mathbb{E}[\gamma(X)g(X)] \geq 1$  formulated in Section 3, where we take  $\gamma(x) = x$  for simplicity. The optimizer  $g_X$  is unique in the case of VaR (Proposition 3). See Section 7 for more details and additional numerical simulations.

problems is described in Section 3. These optimization problems are analyzed in Sections 4, 5 and 6 for VaR, convex risk measures, and expected loss (and utility) functions, respectively. Robustness statements are obtained under general conditions, and further analytical solutions are available in some special cases. As the main message of our results, we see that, for the case of VaR which is argued by many as a robust risk measure, its corresponding optimization is highly non-robust and a small model uncertainty would ruin the optimality of the optimized positions. In sharp contrast, for many convex risk measures including ES and expected loss functions, the optimized positions are generally robust. Some simulation results, in addition to those underlying Figure 1, are provided in Section 7. In Section 8, we present some discussions on the implications of our results for the desirability of specific regulatory risk measures, an ongoing debate in the financial industry (BCBS (2016), IAIS (2014)). Our results yield a (further) strong argument against using VaR as a risk measure within banking and insurance regulation; for a related discussion on robustness in the realm of risk sharing, see Embrechts et al. (2018). In the last section, Section 9, we discuss our notion of robustness in the context of distributionally robust optimization (e.g. Natarajan et al. (2008), Goh and Sim (2010)). The proofs of all results are put in the appendix.

As Goodhart’s law (Goodhart (1984)) implies, when a risk measure becomes a target, it ceases to be a good risk measure.<sup>3</sup> Nevertheless, to what extent this law applies depends on the

<sup>3</sup>Danielsson (2002) applied Goodhart’s law to risk models by saying that a risk model breaks down when used for regulatory purposes.

specific application and each particular risk measure. Our results in this paper thus provide a quantitative and comparative analysis of Goodhart’s law in the context of financial regulation and optimization. Combining our negative result on VaR, which is an example of Goodhart’s law, and our positive result for convex risk measures, our main message may be summarized as:

*As regulatory target, all risk measures cease to be good, but some risk measures, VaR in particular, are much worse than the others.*

## 2 Theoretical framework

### 2.1 Notation

We work with an atomless probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $L^q$  be the set of all random variables in  $(\Omega, \mathcal{F}, \mathbb{P})$  with finite  $q$ -th moment,  $q \in (0, \infty)$ ,  $L^0$  be the space of all  $\mathbb{P}$ -a.s. finite random variables, and  $L^\infty$  the set of all essentially bounded elements of  $L^0$ . For a positive integer  $n$ , write  $L_n^q = (L^q)^n$ . For a vector  $x \in \mathbb{R}^n$ ,  $|x|$  is its Euclidean norm. Throughout, for any  $X \in L^0$ ,  $F_X$  represents the distribution function of  $X$ . The mappings  $\text{ess-inf}(\cdot)$  and  $\text{ess-sup}(\cdot)$  on  $L^0$  stand for the essential infimum and the essential supremum of a random variable, respectively. We write  $X \stackrel{d}{=} Y$  if the random variables  $X$  and  $Y$  have the same distribution under  $\mathbb{P}$ . For  $x \in \mathbb{R}$ , denote by  $\delta_x$  the point-mass probability measure at  $x$ . For real numbers or functions  $x$  and  $y$ , we write  $x \wedge y = \min\{x, y\}$ ,  $x \vee y = \max\{x, y\}$ ,  $x_+ = x \vee 0$ , and  $x_- = (-x) \vee 0$ .

### 2.2 Basic setup of optimization problems

In this section, we first lay out the basic setup for optimization problems when the relevant information on the underlying economic model is known precisely, that is, the case without model uncertainty. Let  $X$  be an  $n$ -dimensional random vector, where  $n$  is a positive integer. The random vector  $X$  is called an *economic vector*, which includes all random sources in an economic model under study, such as potential losses, traded securities, hedging instruments, insurance contracts, macro economic factors, or pricing densities.

Let  $\mathcal{G}_n$  be the set of measurable functions mapping  $\mathbb{R}^n$  to  $\mathbb{R}$ . A random variable  $g(X)$  where  $g \in \mathcal{G}_n$  represents a risky position of an investor, in which positive values represent a loss and negative values represent a gain. This sign convention is in line with the regulatory angle we follow in several applications in our paper. The investor’s problem is to choose among admissible positions  $g(X)$  for some functions  $g$  in an admissible set  $\mathcal{G} \subset \mathcal{G}_n$ .

For a set  $\mathcal{G} \subset \mathcal{G}_n$ , we formulate the problem

$$\text{to minimize: } \rho(g(X)) \quad \text{subject to } g \in \mathcal{G}, \quad (1)$$

where  $\rho$  is an objective functional mapping a set containing  $\{g(X) : g \in \mathcal{G}\}$  to  $\mathbb{R} \cup \{+\infty\}$ . Here one prefers a smaller value of the objective functional over a larger value. Objective functionals considered may be general; examples include (up to a sign change) mean-variance functionals, expected utilities, rank-dependent utility functionals, functionals in cumulative prospect theory, and various risk measures as discussed in Artzner et al. (1999) and Föllmer and Schied (2016). Our main interest will be, however, in the risk measures Value-at-Risk (VaR) and Expected Shortfall (ES<sup>4</sup>).

The elements in the optimization problem (1) can be summarized by an objective functional  $\rho$  and a pair  $(X, \mathcal{G})$ . We always assume that the domain of the objective functional  $\rho$  contains  $\{g(X) : g \in \mathcal{G}\}$ , otherwise (1) is meaningless.

**Example 1.** An illustrative example is the classic problem of hedging in a financial market. Suppose that an investor currently faces a risk  $W$  and would like to hedge against  $W$ . She has access to hedging instruments in a set  $\{g(Y) : g \in \mathcal{G}'\}$  where  $Y$  is an  $(n-1)$ -dimensional economic vector and, say,  $\mathcal{G}' \subset \mathcal{G}_{n-1}$ ,  $n \geq 2$ . Typically, the set  $\mathcal{G}'$  involves a budget constraint. Equivalently, she chooses risky positions in the set  $\{W - g(Y) : g \in \mathcal{G}'\}$ , which represents all possible hedged positions she may attain. Note that  $\{W - g(Y) : g \in \mathcal{G}'\} = \{f(X) : f \in \mathcal{G}\}$  where  $X = (W, Y)$  is an  $n$ -dimensional random vector and

$$\mathcal{G} = \{f \in \mathcal{G}_n : f(w, y) = w - g(y), g \in \mathcal{G}', w \in \mathbb{R}, y \in \mathbb{R}^{n-1}\};$$

therefore the hedging problem is a special case of our general setting (1). Here we allow  $W$  to be arbitrarily dependent on  $Y$ . If  $W$  is a financial risk and  $Y$  is the vector of asset prices in a complete financial market, then it may be that  $W$  is a function of  $Y$ . On the other hand, if  $W$  represents a property and casualty insurance risk and  $Y$  is the vector of asset prices in a financial market, then it may be reasonable to assume that  $W$  and  $Y$  are independent.

**Example 2.** In Markowitz's portfolio selection, an investor wishes to find an optimal allocation vector  $w \in \mathbb{R}^n$  based on a return vector  $Y \in L_n^2$  for  $n$  stocks. The problem can be described as the minimization of  $\mathbb{E}[g(Y)]$  where  $g \in \mathcal{G}'$  and  $\mathcal{G}'$  consists of all functions  $g \in \mathcal{G}_n$  that are of the

---

<sup>4</sup>A formal definition of VaR and ES is given in Section 3. ES is also known as CVaR, CTE, AVaR and TVaR, depending on the context (see e.g. Pflug and Römisch (2007), McNeil et al. (2015) and Föllmer and Schied (2016)). We use the term ES to be consistent with the Basel Committee on Banking Supervision (BCBS (2016)), because our study is motivated by the comparative advantages of VaR and ES in regulation; see discussions in Section 8.

form  $g(y) = \lambda(w^\top y)^2 - w^\top y$  for  $w \in \mathbb{R}^n$  satisfying  $\sum_{i=1}^n w_i = x_0$ . Here,  $\lambda > 0$  is a risk-aversion parameter and  $x_0$  represents the budget constraint of the investor.

In this section, we keep the choice of  $(X, \mathcal{G})$  as generic as possible. Special cases and examples are studied in Sections 3-6.

### 2.3 Uncertainty and robustness against optimization

We proceed to put uncertainty into the optimization problem (1) described above. For  $X \in L_n^0$ ,  $\mathcal{G} \subset \mathcal{G}_n$  and an objective functional  $\rho$ , denote by  $\rho(X; \mathcal{G})$  the minimum possible value of  $\rho$ , namely,

$$\rho(X; \mathcal{G}) = \inf\{\rho(g(X)) : g \in \mathcal{G}\},$$

and by  $\mathcal{G}_X(\rho)$  the set of *optimizing functions*, that is,

$$\mathcal{G}_X(\rho) = \arg \min_{g \in \mathcal{G}} \rho(g(X)) = \{g \in \mathcal{G} : \rho(g(X)) = \rho(X; \mathcal{G})\}. \quad (2)$$

Note that  $\mathcal{G}_X(\rho)$  might be an empty set. Throughout this paper, the notation  $g_X$  will refer to a generic element  $g_X \in \mathcal{G}_X(\rho)$ , and  $g_X(X)$  will be called an *optimized position*.

We shall use  $X$  to represent our (perceived) model for the underlying economic vector. In practice, the model  $X$  is obtained based on stochastic assumptions and statistical inference, and it may not represent a true model for the underlying economic vector. In other words, the optimization problem (1) is often subject to severe model uncertainty. To reflect this issue, let  $\mathcal{Z} \subset L_n^0$  be a set of *possible* economic vectors including  $X$ ;  $\mathcal{Z}$  may be interpreted as the set of *alternative* models.<sup>5</sup> Suppose that the *real* economic vector  $Z \in \mathcal{Z}$  is different from the *perceived* economic vector  $X$ . The information we have at hand is about  $X$  rather than  $Z$ , and we shall refer to  $X$  as the *best-of-knowledge model* and  $Z$  as the *true model*, which is unknowable. We have to make decisions according to the best of our knowledge, that is, as in (2), choosing  $g_X \in \mathcal{G}_X(\rho)$  optimizing our objective  $\rho$ . The *real* but unknown position  $g_X(Z)$  may be different from the *perceived* optimal position  $g_X(X)$ . If  $Z$  and  $X$  are close to each other according to some (pseudo-)metric  $\pi$  (e.g.  $L^\infty$ -metric on the space of bounded random vectors), we would like  $\rho(g_X(Z))$  to be close to  $\rho(g_X(X))$  in order to make sense of the position  $g_X(Z)$ , which may no longer be optimal. In other words, we naturally would desire some continuity of the mapping  $Y \mapsto \rho(g_X(Y))$  at  $Y = X$ .

---

<sup>5</sup>For instance,  $\mathcal{Z}$  can be a parametric family of risk models and hence corresponds to parameter uncertainty of the underlying risk models. Each component of the vector  $X$  may have a different economic meaning. Some of them may be subject to more severe model uncertainty whereas others may be free of model uncertainty. This can be reflected in the choice of  $\mathcal{Z}$  which may be contained in a low-dimensional subset of the set of  $n$ -random vectors.

Note that, in our situation, there is no point in analyzing the problem of optimizing  $g(Z)$  over  $g \in \mathcal{G}$ , because  $Z$  is unknown. This makes our framework conceptually different from the stream of research on stability of the set of optimizers under model uncertainty. In Section 8.2, using examples from financial regulation, we make explicit this very important distinction between our paper and other approaches in the literature.

Putting this into the hedging context of Example 1, suppose that the real economic vector  $Z \in \mathcal{Z}$  is different from  $X$  and an investor has the real risk  $W = h(Z)$  to hedge. The information she has is about  $X$ , and she hedges  $W$  by choosing  $g_X$  from a set of available instruments  $\mathcal{G}'$ . In this case,  $h(Z) - g_X(Z)$  is the remaining risk she actually faces after hedging. Under this setting, assuming the quality of model  $X$  is good,  $Z$  should be close to  $X$  in some sense, and she would naturally desire some continuity of the function  $Y \mapsto \rho(h(Y) - g_X(Y))$  at  $Y = X$ .

The admissible set  $\mathcal{G}$  is not subject to model uncertainty, as the investor knows which positions she can choose in the optimization problem. For instance, in the above hedging example, a budget constraint that determines  $\mathcal{G}$  is not affected by the model assumptions made for  $X$ ; it is simply the observed prices for the hedging instruments.

In light of the above consideration, we endow the set  $\mathcal{Z}$  of all possible economic vectors with a pseudo-metric  $\pi$ . Common choices of  $\pi$  are the  $L^\infty$ , the  $L^q$ , the Wasserstein, and the weak (pseudo-)metrics; see Example 3 below. The reason for considering a pseudo-metric instead of a metric is to be able to incorporate objective functionals based on the distributions of risks, for instance, law-invariant risk measures and expected utility functions.<sup>6</sup>

**Definition 1.** We call  $(\mathcal{G}, \mathcal{Z}, \pi)$  an *uncertainty triplet* if  $\mathcal{G} \subset \mathcal{G}_n$  and  $(\mathcal{Z}, \pi)$  is a pseudo-metric space of  $n$ -random vectors. For a given uncertainty triplet  $(\mathcal{G}, \mathcal{Z}, \pi)$ , we say that an objective functional  $\rho$  is *compatible* if  $\rho$  maps  $\mathcal{G}(\mathcal{Z}) = \{g(Z) : Z \in \mathcal{Z}, g \in \mathcal{G}\}$  to  $\mathbb{R} \cup \{+\infty\}$ , and  $\rho(g(Y)) = \rho(g(Z))$  for all  $g \in \mathcal{G}$  and  $Y, Z \in \mathcal{Z}$  with  $\pi(Y, Z) = 0$ , i.e.  $Y$  and  $Z$  are indistinguishable under the pseudo-metric  $\pi$ .

**Definition 2.** Let  $(\mathcal{G}, \mathcal{Z}, \pi)$  be an uncertainty triplet. A compatible objective functional  $\rho$  is *robust against optimization at  $X \in \mathcal{Z}$  for  $(\mathcal{G}, \mathcal{Z}, \pi)$*  if there exists  $g_X \in \mathcal{G}_X(\rho)$  such that the function  $Y \mapsto \rho(g_X(Y))$  is  $\pi$ -continuous at  $Y = X$ .

In this paper, we are mainly interested in robustness in the sense of Definition 2, and it should not be confused with the classic qualitative robustness of risk measures as studied in e.g. Cont et al. (2010), Kou et al. (2013), Krättschmer et al. (2014) and Embrechts et al. (2015).

---

<sup>6</sup>Because of the extensive use of simulation and estimation methods for risk evaluation, uncertainty at the level of distributions is the most common in risk management practice; on the other hand, in optimization problems, risks with identical distributions are not equivalent (see our functional optimization problems in Section 3). For this reason, we do not use the equivalent class induced by the pseudo-metric.



On the other hand, in contrast to the robust optimization literature (e.g. [Goh and Sim \(2010\)](#), [Wiesemann et al. \(2014\)](#)), our focus is the robustness of objective functionals in optimization, instead of *how to solve* particular optimization problems. As such, our setup and methodology are also different from classic ones in the optimization literature. Section 8.2 contains further discussions on the formulation of Definition 2, including some possible alternatives.

Below we give three prominent examples of  $\pi$ , which will appear throughout the paper.

**Example 3.** (i) For a subset  $\mathcal{Z}$  of  $L_n^\infty$ , the  $L^\infty$ -metric  $\pi_n^\infty$  is defined as

$$\pi_n^\infty(X, Y) = \|X - Y\|_\infty = \text{ess-sup}(|X - Y|), \quad X, Y \in \mathcal{Z}. \quad (3)$$

(ii) For  $q \in [1, \infty)$  and a subset  $\mathcal{Z}$  of  $L_n^q$ , the  $L^q$ -metric  $\pi_n^q$  is defined as

$$\pi_n^q(X, Y) = \|X - Y\|_q = (\mathbb{E}[|X - Y|^q])^{\frac{1}{q}}, \quad X, Y \in \mathcal{Z}. \quad (4)$$

(iii) For a subset  $\mathcal{Z}$  of  $L_n^0$ , the pseudo-metric  $\pi_n^W$  is defined as

$$\pi_n^W(X, Y) = \pi_P(F_X, F_Y), \quad X, Y \in \mathcal{Z}. \quad (5)$$

where  $\pi_P$  is the Prokhorov metric over the set of probability distribution measures.<sup>7</sup> In this case, convergence in  $\pi_n^W$  is equivalent to convergence in distribution (or weak convergence). One could also replace  $\pi_P$  in (5) by a Wasserstein metric over probability measures, and obtain a Wasserstein pseudo-metric on a suitable subspace of  $L_n^0$ .

## 2.4 Basic properties on robustness and continuity

We first explain a few basic properties about Definition 2. Robustness against optimization is a joint property of  $(\rho, X, \mathcal{G}, \mathcal{Z}, \pi)$ , and only a  $\pi$ -neighbourhood of  $X$  in  $\mathcal{Z}$  matters in the definition. If  $\rho$  is robust against optimization at  $X$  for  $(\mathcal{G}, \mathcal{Z}, \pi)$ , then  $\rho$  is also robust against optimization at  $X$  for  $(\mathcal{G}, \mathcal{Z}', \pi)$  if  $X \in \mathcal{Z}' \subset \mathcal{Z}$ , and the same holds true for  $(\mathcal{G}, \mathcal{Z}, \pi')$  if  $\pi'$  is a stronger pseudo-metric than  $\pi$ . On the other hand, if the optimization problem does not admit a solution, that is,  $\mathcal{G}_X(\rho) = \emptyset$ , then  $\rho$  is not robust against optimization at  $X$ .

Robustness of  $\rho$  relies on both some continuity of  $\rho$  on  $\mathcal{G}(\mathcal{Z})$  and some continuity of functions in  $\mathcal{G}_X(\rho)$ . In what follows, we give a few general results where  $\mathcal{G}_X(\rho)$  contains a continuous function. Whereas these results are fairly simple, they could nevertheless be useful in situations

<sup>7</sup>Precisely,  $\pi_P(\mu, \nu) = \inf \{ \varepsilon > 0 : \mu(A) \leq \nu(A_\varepsilon) + \varepsilon \text{ and } \nu(A) \leq \mu(A_\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{B}(\mathbb{R}^n) \}$ , where  $A_\varepsilon = \{x \in \mathbb{R}^n : \|x - y\| < \varepsilon \text{ for some } y \in A\}$  and  $\|\cdot\|$  is the Euclidean norm.

where  $\mathcal{G}$  is nice enough. However, later we will see that in many representative problems,  $\mathcal{G}_X(\rho)$  does not necessarily have any continuous elements for commonly used risk measures such as VaR and convex risk measures; their robustness properties will be investigated in Sections 4 and 5.

For a bijection  $g \in \mathcal{G}_n$  and a pseudo-metric space  $(\mathcal{Z}, \pi)$  of  $n$ -dimensional random vectors, let  $(g(\mathcal{Z}), \pi_g)$  be another pseudo-metric space defined as

$$\pi_g(g(X), g(Y)) = \pi(X, Y) \quad \text{for } X, Y \in \mathcal{Z}.$$

**Proposition 1.** *Suppose that for an uncertainty triplet  $(\mathcal{G}, \mathcal{Z}, \pi)$ ,  $X \in \mathcal{Z}$  and a compatible objective functional  $\rho$ ,  $\mathcal{G}_X(\rho)$  contains a bijection  $g$ , and  $\rho$  is  $\pi_g$ -continuous on  $g(\mathcal{Z})$ . Then  $\rho$  is robust against optimization at  $X$  for  $(\mathcal{G}, \mathcal{Z}, \pi)$ .*

Next we look at the basic settings of  $\mathcal{Z} = L_n^\infty$ ,  $\mathcal{Z} = L_n^q$ , and  $\mathcal{Z} = L_n^0$ , equipped with the  $L^\infty$  metric  $\pi_n^\infty$  in (3), the  $L^q$  metric  $\pi_n^q$  in (4), and the pseudo-metric  $\pi_n^W$  in (5), respectively.

**Proposition 2.** *Let  $\rho$  be a compatible objective functional for the uncertainty triplet  $(\mathcal{G}, \mathcal{Z}, \pi)$  and  $X \in \mathcal{Z}$ .*

- (i) *Suppose  $(\mathcal{Z}, \pi) = (L_n^\infty, \pi_n^\infty)$ . If  $\mathcal{G}_X(\rho)$  contains a continuous function  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $\rho$  is  $\pi_1^\infty$ -continuous, then  $\rho$  is robust against optimization at  $X$  for  $(\mathcal{G}, \mathcal{Z}, \pi)$ .*
- (ii) *Suppose  $(\mathcal{Z}, \pi) = (L_n^q, \pi_n^q)$ ,  $q \in [1, \infty)$ . If  $\mathcal{G}_X(\rho)$  contains a continuous and linearly growing<sup>8</sup> function  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $\rho$  is  $\pi_1^q$ -continuous, then  $\rho$  is robust against optimization at  $X$  for  $(\mathcal{G}, \mathcal{Z}, \pi)$ .*
- (iii) *Suppose  $(\mathcal{Z}, \pi) = (L_n^0, \pi_n^W)$ . If  $\mathcal{G}_X(\rho)$  contains a continuous function  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $\rho$  is  $\pi_1^W$ -continuous, then  $\rho$  is robust against optimization at  $X$  for  $(\mathcal{G}, \mathcal{Z}, \pi)$ .*

Proposition 2 provides simple criteria for verifying robustness of some objective functionals based on continuity of the optimizing functions in  $\mathcal{G}_X(\rho)$ . As we shall see in Sections 4-5, for the popular risk measures VaR and ES, such criteria may not be very useful, as typically the optimizing functions lack the corresponding continuity. More detailed analyses are needed to draw meaningful conclusions for these objectives, which will be the focus of the next few sections.

### 3 A class of functional optimization problems

Our main interest is in the robustness of risk measures in optimization, and in particular, Value-at-Risk (VaR) and Expected Shortfall (ES). Here, a positive value of  $Y$  represents a loss

---

<sup>8</sup>A function  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  is *linearly growing* if for some  $C > 0$ ,  $|g(y)| \leq C|y|$  for all  $y \in \mathbb{R}^n$  with  $|y| > 1$ . This property is satisfied by, for instance, Lipschitz-continuous functions.

and a negative value represents a gain. The VaR at confidence level  $p \in (0, 1)$  is defined as

$$\text{VaR}_p(Y) = \inf\{x \in \mathbb{R} : \mathbb{P}(Y \leq x) \geq p\} = F_Y^{-1}(p), \quad Y \in L^0,$$

and the ES at confidence level  $p \in (0, 1)$  is defined as

$$\text{ES}_p(Y) = \frac{1}{1-p} \int_p^1 \text{VaR}_s(Y) ds, \quad Y \in L^0. \quad (6)$$

Note that  $\text{ES}_p(Y)$  may take the value  $\infty$  if  $Y$  is not integrable. In addition, we write

$$\text{ES}_1(Y) = \text{VaR}_1(Y) = \text{ess-sup}(Y) = \sup\{x \in \mathbb{R} : \mathbb{P}(Y \leq x) < 1\}.$$

We summarize some well known robustness properties of  $\text{VaR}_p$  and  $\text{ES}_p$  for  $p \in (0, 1)$  below.

- 1.)  $\text{VaR}_p$  is continuous with respect to convergence in distribution if and only if the inverse cdf of  $X$  is continuous at  $p$ ; see, e.g., Proposition 7.3.1 in [Shorack \(2000\)](#). Note that convergence in distribution is weaker than (Wasserstein)  $L^q$ -convergence for  $q \in [1, \infty]$ .
- 2.) It follows immediately from 1.) and (6) that  $\text{ES}_p$  is continuous with respect to convergence in distribution on every uniformly integrable subset of  $L^1$ . In particular,  $\text{ES}_p$  is continuous with respect to  $L^q$ -convergence for  $q \in [1, \infty]$ . On the other hand,  $\text{ES}_p$  is not continuous with respect to convergence in distribution on any set containing  $L_+^\infty$ .

In addition to VaR and ES, we will consider two general classes of convex risk measures, as well as expected utility and loss functions, which will be introduced in Sections 5 and 6.

Next, we describe a general class of functional optimization problems. For an  $n$ -dimensional random vector  $X$ , two measurable functions  $v : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{-\infty\}$  and  $w : \mathbb{R}^n \rightarrow \mathbb{R}$ , a measurable price density  $\gamma : \mathbb{R}^n \rightarrow (0, \infty)$ , and a constant  $x_0 \in \mathbb{R}$ , we consider the following set

$$\mathcal{G} = \left\{ g \in \mathcal{G}_n : v \leq g \leq w \text{ and } \mathbb{E}[g(X)\gamma(X)] \geq x_0 \right\}. \quad (7)$$

The corresponding optimization problem is

$$\text{to minimize: } \rho(g(X)) \quad \text{subject to } v \leq g \leq w, \quad \mathbb{E}[\gamma(X)g(X)] \geq x_0. \quad (8)$$

The optimization problem (8) is called ‘functional’ because the objective is optimized over a large set of  $n$ -variate functions.

Intuitively, the functions  $v$  and  $w$  describe lower and upper bounds on admissible functions

$g \in \mathcal{G}$ , while the condition  $\mathbb{E}[g(X)\gamma(X)] \geq x_0$  describes a budget constraint.<sup>9</sup> Hence, the problem (8) represents portfolio optimization with given budget, which is a classic problem in quantitative finance, and includes many interesting special cases. Below we present two simple examples, one in the context of a complete financial market and the other one in the context of insurance design. Since this problem has attracted substantial interest in its own right, we will en passant contribute to the corresponding literature; see Remark 2.

**Example 4** (Optimal investment). The optimization problem (8) connects as follows with continuous-time optimal hedging problems in a complete market with an arbitrary number of primary assets. Suppose that  $S = (S_t)_{t \in [0, T]}$  is a  $d$ -dimensional semimartingale admitting a unique local martingale measure  $\mathbb{Q}$  with price density  $\gamma = d\mathbb{Q}/d\mathbb{P}$  on  $\mathcal{F} = \mathcal{F}_T^S$ . We can interpret  $S$  as the discounted price process of  $d$  risky securities. The fact that  $\mathbb{Q}$  is unique is equivalent to the completeness of the model. Let  $X$  be a random vector with  $\sigma(X) = \mathcal{F}_T^S$ , which represents market randomness (such a random vector exists under mild conditions such as continuity of  $S$ ). In this context, for an investor who needs to pay a random wealth  $f(X)$  at time  $T$ , it is a natural task to minimize, e.g.,  $\rho(f(X) - V_T)$ , where  $V_T := V_T(X)$  is the discounted time- $T$  value of a self-financing trading strategy satisfying a cost constraint on the initial investment  $V_0$  and certain other constraints such as  $v(X) \leq V_T \leq w(X)$  for some functions  $v$  and  $w$ . By martingale arguments, the initial investment satisfies  $V_0 = \mathbb{E}[\gamma(X)V_T]$ . On the other hand, market completeness implies that  $S$  has the martingale representation property, and so every feasible functional  $g(X)$  can be represented in the form  $f(X) - V_T$ , where  $V$  is again the value process of some self-financing trading strategy. So, by letting  $g(x) := f(x) - V_T(x)$ , we arrive at a special case of (7).

**Example 5** (Insurance design). The problems of insurance design, pioneered by Arrow (1963), can also be described by (8). Let  $X \geq 0$  represent a random future loss to an insured, and  $f(x)$  represent the amount of payment from the insurer if the realized loss  $X$  is equal to  $x$ ;  $f$  is called an insurance indemnity function. Suppose  $\gamma \geq 1$  is a constant, and  $\gamma\mathbb{E}[f(X)]$  is used to price the insurance contract with payment  $f$  (premium based on the expected payment is called the actuarial premium). Let  $y_0$  be the budget of the insured. The standard optimal insurance problem of the insured with risk measure  $\rho$  is

$$\text{to minimize: } \rho(X - f(X)) \quad \text{subject to } 0 \leq f(X) \leq X, \quad \gamma\mathbb{E}[f(X)] \leq y_0,$$

which belongs to (8) by choosing  $g(x) = x - f(x)$ ,  $w(x) = x$ ,  $v(x) = 0$ , and  $x_0 = \gamma\mathbb{E}[X] - y_0$ .

---

<sup>9</sup>Using our sign convention, by holding a risky position  $g(X)$ , one receives the monetary amount  $\mathbb{E}[\gamma(X)g(X)]$ . Equivalently, one pays  $-\mathbb{E}[\gamma(X)g(X)] \leq -x_0$ , thus the usual budget constraint.

Some other requirements on the payment  $f$  may be further imposed; see e.g., [Bernard et al. \(2015\)](#) and the references therein.

To study the robustness of risk measures for this problem, the main assumption on  $(\mathcal{G}, \mathcal{Z}, \pi)$  and  $X$  is given below, which will be assumed in the next three sections. The assumption on  $X$  is standard and satisfied by practically all financial models in which the assets prices have densities.

**Assumption G.**  $\mathcal{G}$  is given by (7) where  $\mathbb{E}[\gamma(X)] < \infty$  and  $\mathcal{G} \neq \emptyset$ ; the distribution measure  $\mu_X$  of  $X$  has a positive density on its support, which is a convex subset of  $\mathbb{R}^n$ , and  $(\mathcal{Z}, \pi)$  is  $(L_n^0, \pi_n^W)$  or  $(L_n^q, \pi_n^q)$ ,  $q \in [1, \infty]$ .

*Remark 1.* In Sections 4-6 below, we will study robustness of VaR, convex risk measures, and utility functions under a few further assumptions on the optimization problem (7). Our main results show that VaR is not robust under a very weak assumption (Assumption V), whereas in sharp contrast, convex risk measures are robust under some standard regularity assumptions (Assumptions P and R).

In general, it is difficult to obtain analytical solutions to (8); instead, we will obtain robustness statements on various risk measures in the subsequent sections. Explicit optimizers can be obtained for a particular case of (8) such that  $X$  is one-dimensional,  $v(x) = 0$  and  $w(x) = x$  (see e.g., Example 5); that is,

$$\text{to minimize: } \rho(g(X)) \quad \text{subject to } 0 \leq g(X) \leq X, \quad \mathbb{E}[\gamma(X)g(X)] \geq x_0. \quad (9)$$

## 4 Robustness of Value-at-Risk

The main task of this section is to establish the (non-)robustness of VaR for the optimization problem (8) in Section 3. We will make the following assumption on  $\gamma$ ,  $v$  and  $w$ , as well as the minimum value of the risk measure  $\rho = \text{VaR}_p$  for  $p \in (0, 1)$ . Recall that  $\rho(X; \mathcal{G}) = \inf\{\rho(g(X)) : g \in \mathcal{G}\}$ .

**Assumption V.**  $\text{ess-sup}(v) < \rho(X; \mathcal{G}) < \rho(w(X))$  and  $\gamma$  is bounded from above.

Assumption V is quite general and weak. The condition  $\text{ess-sup}(v) < \rho(X; \mathcal{G})$ , meaning that the lower bound  $v$  is not too large, is primarily used to guarantee that the budget constraint is binding. The condition  $\rho(X; \mathcal{G}) < \rho(w(X))$  simply says that the optimization problem is not solved by trivially choosing the largest possible position  $g = w$ . The boundedness condition on  $\gamma$  may be significantly relaxed as we only need boundedness in a neighbourhood of one specific

point. An explanation and a technical discussion on these conditions are put in Remark 4 in Appendix A.2.

**Theorem 1.** *For  $p \in (0, 1)$ , under Assumptions G and V,  $\rho = \text{VaR}_p$  is not robust against optimization at  $X$  for  $(\mathcal{G}, \mathcal{Z}, \pi)$ .*

Theorem 1 implies that, for the optimization problem (8) and all choices of commonly used (pseudo-)metrics,  $\text{VaR}_p$  is not robust against optimization, and this result holds for a general continuously distributed random vector  $X$ . As a consequence,  $\text{VaR}_p$  has the poorest possible robustness in our setup. The main reason behind this phenomenon is quite intuitive: as a key point in the proof of Theorem 1, any optimizing function  $g_X$  always has a jump at the  $p$ -quantile of  $g_X(X)$ , making it most vulnerable to model uncertainty.

In practice, one may consider a subset  $\mathcal{G}'$  of  $\mathcal{G}$  which contains only continuous functions, so that robustness holds by Proposition 2. We emphasize that the optimization for VaR is still problematic in this setting: One needs to search for functions in  $\mathcal{G}'$  which closely approximate the discontinuous function  $g_X$ , and hence the stability of the optimization is still weak.

Next, we consider the one-dimensional setting (9), which corresponds to  $v(x) = 0$  and  $w(x) = x$ . Denote by  $q := \text{VaR}_p(X; \mathcal{G})$  the minimum value of (9) with  $\rho = \text{VaR}_p$ , and we have  $q > 0$  by Assumption V. The next proposition gives an explicit solution to (9), under a further minor condition that  $(X - q)\gamma(X)$  has a unique  $p$ -quantile.

**Proposition 3.** *Let  $p \in (0, 1)$ ,  $\rho = \text{VaR}_p$  and  $Y = (X - q)\gamma(X)$ . Suppose that Assumptions G and V hold,  $\mathbb{E}[\gamma(X)X] < \infty$ , and  $\mathbb{P}(Y \leq \text{VaR}_p(Y)) = p$ . Problem (9) admits a  $\mu_X$ -a.s. unique solution that is of the form*

$$g_X(x) = x\mathbb{1}_{\{(x-q)\gamma(x) > c\}} + (x \wedge q)\mathbb{1}_{\{(x-q)\gamma(x) \leq c\}}, \quad (10)$$

where  $c = \text{VaR}_p(Y)$ . Moreover,

$$p\text{ES}_{1-p}(-Y_+) = x_0 - \mathbb{E}[\gamma(X)X]. \quad (11)$$

Since  $Y_+ = \gamma(X)(X - q)_+$ , the left-hand side of (11) is an increasing function of  $q$ , and hence the value  $q$  can be numerically computed by solving (11). Section 7 contains simulation studies for the problem setting in Proposition 3.

## 5 Robustness of two classes of convex risk measures

In this section, we obtain positive robustness results for two important classes of convex risk measures, the *divergence risk measures* and the *utility-based shortfall risk measures*. The first class of risk measures, which are sometimes also called optimized certainty equivalents, contains Expected Shortfall as an important special case. The second class comprises the expectiles. Both classes intersect at the entropic risk measure, which can also be treated within the setting of our subsequent Section 6, where we will analyze the robustness of expected utility/expected loss.

We continue to study the optimization problem (8) with the following regularity conditions.

**Assumption P.** The price density  $\gamma : \mathbb{R}^n \rightarrow (0, \infty)$  is  $\mu_X$ -a.e. continuous and  $\gamma(X)$  has a continuous density.

**Assumption R.** The functions  $v$  and  $w$  are  $\mu_X$ -a.e. continuous. Moreover,  $-\infty \leq \mathbb{E}[\gamma(X)v(X)] \leq x_0 \leq \mathbb{E}[\gamma(X)w(X)] \leq \mathbb{E}[|\gamma(X)w(X)|] < \infty$ .

Assumptions P and R are simple continuity and integrability requirements on  $\gamma$ ,  $v$  and  $w$ , and they are satisfied in, for instance, a multi-dimensional Black-Scholes market model in which both  $v(X)$  and  $w(X)$  have finite prices.

We consider a convex risk measure, i.e., a functional  $\rho : L^1 \rightarrow \mathbb{R} \cup \{+\infty\}$  satisfying monotonicity, cash invariance, and convexity (see, e.g., Chapter 4 in Föllmer and Schied (2016)). Let  $\varphi : \mathbb{R} \rightarrow [0, +\infty]$  be a proper closed convex function whose effective domain is an interval with endpoints  $a < b$ . We assume moreover that  $a < 1 < b$  and that  $0 = \varphi(1) = \min_x \varphi(x)$ . Then the  $\varphi$ -divergence of a probability measure  $\mathbb{Q}$  with respect to  $\mathbb{P}$  is

$$I_\varphi(\mathbb{Q}|\mathbb{P}) := \begin{cases} \int \varphi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right) d\mathbb{P} & \text{if } \mathbb{Q} \ll \mathbb{P}, \\ +\infty & \text{otherwise.} \end{cases} \quad (12)$$

The corresponding divergence risk measure is defined as

$$\rho(Y) := \sup_{\mathbb{Q} \ll \mathbb{P}} (\mathbb{E}_{\mathbb{Q}}[Y] - I_\varphi(\mathbb{Q}|\mathbb{P})), \quad Y \in L^\infty. \quad (13)$$

If  $\varphi(x) = x \log x - x + 1$ , then  $I_\varphi(\mathbb{Q}|\mathbb{P})$  is the relative entropy, or the Kullback–Leibler divergence, of  $\mathbb{Q}$  with respect to  $\mathbb{P}$  and  $\rho$  is an entropic risk measure. If  $\varphi = \infty \cdot \mathbf{1}_{[1/(1-p), \infty)}$  for some  $p \in [0, 1)$ , then  $\rho$  is the Expected Shortfall,  $\text{ES}_p$ .

**Theorem 2.** *In addition to Assumptions G, P, and R, we assume that  $v$  and  $w$  are bounded. Then the divergence risk measure  $\rho$  is robust against optimization at  $X \in L_n^0$  for  $(\mathcal{G}, L_n^0, \pi_n^W)$ .*

The proof of Theorem 2 relies on the duality formula from Ben-Tal and Teboulle (1987, 2007), which for general divergence risk measures has only been established on  $L^\infty$ . This is one of the reasons for assuming the boundedness of  $v$  and  $w$  in Theorem 2. It is possible, however, to relax their boundedness by imposing a suitable growth condition, given that the aforementioned duality formula extends to  $L^p$  (or, more generally, to a certain Orlicz space). An important special case for which this is possible is Expected Shortfall,  $\text{ES}_p$ . Recall that the Expected Shortfall has the following representation (e.g., McNeil et al. (2015, Theorem 8.14)) for  $Y \in L^1$ ,

$$\text{ES}_p(Y) = \frac{1}{1-p} \int_p^1 F_Y^{-1}(s) ds = \sup \left\{ \mathbb{E}_{\mathbb{Q}}[Y] : \mathbb{Q} \ll \mathbb{P} \text{ and } \frac{d\mathbb{Q}}{d\mathbb{P}} \leq \frac{1}{1-p} \right\}. \quad (14)$$

Note that the right-hand representation coincides with (13) if  $Y \in L^\infty$  and  $\varphi = \infty \cdot \mathbb{1}_{[1/(1-p), \infty)}$ . We will say that a function  $f : \mathbb{R}^k \rightarrow \mathbb{R}$  has growth index  $q \in [0, \infty]$ , if  $f$  is locally bounded for  $q = \infty$  and if for  $q < \infty$  there exists a constant  $c$  such that  $|f(x)| \leq c(1 + |x|^q)$  for  $x \in \mathbb{R}^k$ .

**Corollary 1.** *In addition to Assumptions G, P, and R, we assume that both  $v$  and  $w$  have growth index  $q \in [1, \infty]$ . Then Expected Shortfall,  $\text{ES}_p$ , with  $p \in (0, 1)$ , is robust against optimization at  $X \in L_n^q$  for  $(\mathcal{G}, L_n^q, \pi_n^q)$ .*

Comparing Theorem 1 and Corollary 1, we see that ES has clear advantages over VaR in terms of robustness against optimization. Note that Assumptions G and V in Section 4 and Assumptions P and R in this section are simple regularity conditions, and they are realistic for practical models. Therefore, we safely can say that for Problem (8), VaR is generally not robust against optimization, and ES is generally robust against optimization.

In Theorems 1-2 and Corollary 1, we assumed that  $w$  is finite, and thus the admissible positions in the optimization problem (8) has an upper bound. A similar comparison on the robustness of VaR and ES is obtained for the unbounded problem ( $w = \infty$  and  $v = -\infty$ ), which is presented in Appendix A.6.

Now we turn to an analysis of the robustness of utility-based shortfall risk measures as introduced in Föllmer and Schied (2002). To this end, let  $\ell : \mathbb{R} \rightarrow \mathbb{R}$  be a nonconstant, increasing, and convex loss function and  $x_0$  be an interior point in the range of  $\ell$ . The corresponding utility-based shortfall risk measure is given by

$$\rho(Y) = \inf \{ m \in \mathbb{R} : \mathbb{E}[\ell(Y - m)] \leq x_0 \}, \quad Y \in L^\infty.$$

**Theorem 3.** *In addition to Assumptions G, P, and R, we assume that  $v$  and  $w$  are bounded. Then the utility-based shortfall risk measure  $\rho$  is robust against optimization at  $X \in L_n^0$  for  $(\mathcal{G}, L_n^0, \pi_n^W)$ .*



A notable special case of a utility-based shortfall risk measure is the expectile of  $Y \in L^1$  at level  $\tau \in [0, 1]$ , defined as the unique solution to the equation

$$\tau \mathbb{E}[(Y - z)_+] = (1 - \tau) \mathbb{E}[(Y - z)_-].$$

Expectiles were introduced by [Newey and Powell \(1987\)](#) and have recently gained attention in the context of the discussion on backtesting risk estimates. As stated<sup>10</sup> in Exercise 4.9.2.(b) of [Föllmer and Schied \(2016\)](#), for  $Y \in L^\infty$  and  $\tau \in (1/2, 1]$ , the expectile is equal to the utility-based shortfall risk for the loss function  $\ell(x) = \tau x_+ - (1 - \tau)x_-$ . For the special case of an expectile, the following corollary relaxes the boundedness condition by a more general growth condition, just as we did for Expected Shortfall in Corollary 1.

**Corollary 2.** *In addition to Assumptions G, P, and R, we assume that both  $v$  and  $w$  have growth index  $q \in [1, \infty]$ . Then the expectile at level  $\tau \in (1/2, 1]$  is robust against optimization at  $X \in L_n^q$  for  $(\mathcal{G}, L_n^q, \pi_n^q)$ .*

*Remark 2.* The results presented in this section rely on obtaining concrete solutions to the problem of minimizing  $\rho(g(X))$  over  $g \in \mathcal{G}$ . For constant constraint functions,  $v$  and  $w$ , and a coherent risk measure  $\rho$ , this problem can be formulated as a composite hypothesis testing problem, and there exists a significant amount of corresponding literature; see, e.g., Sections 3.5, 8.3 and the corresponding bibliographical notes in [Föllmer and Schied \(2016\)](#) for a summary. Much fewer results were obtained for the case of nonconstant constraint functions, and those that are available often lack some concreteness. A notable exception is [Sekine \(2004\)](#) in a one-dimensional setting which solves (9); see Proposition 4 below. It is therefore worth pointing out that our proofs also provide structure results for the solutions of our optimization problems. Specifically, in the context of Corollary 1, our proof yields that there exists a minimizer  $g_X$  that has one of the following two forms, where  $z^* \in \mathbb{R}$  and  $c > 0$  are suitable constants:

$$g_X(x) = \begin{cases} (v(x) \vee z^* \wedge w(x)) \mathbb{1}_{\{0 < c\gamma(x) < 1\}} & \text{or} \\ (v(x) \vee z^* \wedge w(x)) \mathbb{1}_{\{c\gamma(x) > 1\}} + v(x) \mathbb{1}_{\{c\gamma(x) \leq 1\}} \end{cases}$$

Moreover, in the settings of Theorem 2 and of Theorem 3, there exists a constant  $z^*$  such that a minimizer  $g_X$  is also a minimizer of the expected loss  $\mathbb{E}[\ell(g(X) - z^*)]$  over  $g \in \mathcal{G}$ , where  $\ell(x) := \sup_{y \geq 0} (xy - \varphi(y))$  for Theorem 2. The problem of minimizing this expected loss and its robustness properties will be discussed in the subsequent Section 6.

Finally, we consider the special setting (9), which corresponds to  $v(x) = 0$  and  $w(x) = x$ ,

---

<sup>10</sup>Note the differing sign convention in [Föllmer and Schied \(2016\)](#).

for  $\rho = \text{ES}_p$ ,  $p \in (0, 1)$ . This problem has an explicit solution based on Theorem 8.26 of [Föllmer and Schied \(2016\)](#), which is a slight generalization of a result by [Sekine \(2004\)](#).

**Proposition 4.** *Let  $p \in (0, 1)$  and  $\rho = \text{ES}_p$ . Suppose that Assumption [P](#) holds and  $0 \leq x_0 < \mathbb{E}[\gamma(X)X]$ . There exist constants  $d > 0$  and  $r \geq 0$  such that the function*

$$g_X(x) = x\mathbb{1}_{\{\gamma(x) > d\}} + (x \wedge r)\mathbb{1}_{\{\gamma(x) \leq d\}}, \quad x \in \mathbb{R}, \quad (15)$$

*solves Problem [\(9\)](#). Moreover,  $r$  is a  $p$ -quantile of  $g_X(X)$ .*

For given  $X$ ,  $\gamma$ , and  $r$ , one can compute  $d$  as a function of  $r$  by numerically solving the equation  $\mathbb{E}[\gamma(X)g_X(X)] = x_0$ . Subsequently, one can find the optimal  $r$  by numerically minimizing the expression  $\text{ES}_p(g_X(X))$ ; see [Section 7](#) for its implementation in simulation studies.

## 6 Robustness of expected utility and loss functions

An expected loss  $\rho_\ell$  is a mapping

$$Y \mapsto \rho_\ell(Y) = \mathbb{E}[\ell(Y)],$$

where  $\ell : \mathbb{R} \rightarrow \mathbb{R}$  is a nonconstant, nondecreasing, and convex function. Up to a sign change and a constant shift, minimizing an expected loss is equivalent to maximizing an expected utility, via the relation  $\ell(x) = -u(b - x)$ , where  $u$  is a concave utility function, and  $b$  is the constant wealth level of the decision maker. As in [Section 5](#), we consider the set  $\mathcal{G}$  in [\(7\)](#). The problem of minimizing an expected loss—or, equivalently, of maximizing expected utility—under a budget constraint has a long history; see, e.g., [Ekeland and Temam \(1976\)](#) or, within a financial context, [Section 3.3 in Föllmer and Schied \(2016\)](#). In the existing results, the loss function is typically assumed to be continuously differentiable and often required to be strictly convex and to satisfy the Inada conditions (e.g., [Föllmer and Schied \(2016, p.160\)](#)). Let us point out that none of these assumptions is imposed here. This level of generality is crucial for us, because our subsequent [Theorem 4](#) forms the basis for the proofs of the results in [Section 5](#). For instance, in [Corollary 1](#), it is applied to the loss function  $\ell(x) = x_+ = 0 \vee x$ , which clearly satisfies none of the classical requirements. In what follows,  $\ell_+$  and  $\ell_-$  are the positive and negative parts of  $\ell$ , respectively.

**Theorem 4.** *Suppose that Assumptions [G](#), [P](#), and [R](#) hold. Let  $\ell_+$  have growth index  $q^+ \in [1, \infty]$  and suppose that  $w$  has growth index  $p \in [1, \infty]$ . If, moreover,  $\ell$  is not bounded from below, let  $\ell_-$  have growth index  $q^- \in [0, 1]$  and suppose that the growth index  $r$  of  $v$  satisfies  $r < \infty$  if  $q^- = 0$  and  $r \leq pq^+/q^-$  otherwise. Then the expected loss  $\rho_\ell$  is robust against optimization at*

$X \in L_n^{pq^+}$  for  $(\mathcal{G}, L_n^{pq^+}, \pi_n^{pq^+})$ .

If  $\ell$  is bounded from below, our assumptions allow us to take  $v \equiv -\infty$ , in which case the lower bound is immaterial. If, on the other hand, both  $v$  and  $w$  are bounded, then the expected loss  $\rho_\ell$  will be robust at  $X \in L_n^0$  for  $(\mathcal{G}, L_n^0, \pi_n^W)$ .

Theorem 4 implies that the optimization of expected loss or expected utility is also generally robust under mild regularity conditions. This is similar to the optimization of convex risk measures (Theorems 2 and 3), and in sharp contrast to that of VaR (Theorem 1).

## 7 Simulation results

In this section, we illustrate the robustness and non-robustness of ES and VaR against optimization by means of numerical simulations based on the formulas obtained in Propositions 3 and 4 for Problem (9). In our setup, the true risk factor  $Z$  is either exponentially or Pareto distributed with an unknown parameter  $\theta$ .<sup>11</sup> The investor obtains an estimate  $\hat{\theta}$  for  $\theta$  and considers a corresponding model  $X$ . Then the investor minimizes  $\rho(g(X))$  within the class of all measurable functions  $g$  satisfying the inequality  $0 \leq g(x) \leq x$  and the budget constraint  $\mathbb{E}[\gamma(X)X] \geq x_0$ . We consider both  $\rho = \text{VaR}_{0.99}$  and  $\rho = \text{ES}_{0.975}$ , where the respective levels 0.99 and 0.975 are chosen according to Basel III regulation (BCBS (2016)). To keep things simple, we let  $\gamma(x) = x$ ; this choice allows us to compute several auxiliary quantities in closed form, thus reducing the possible impact of numerical errors.

In Figure 1, we have seen that, in the case of Pareto-distributed risks, the true risk  $\text{VaR}_{0.99}(g_X(Z))$  is substantially larger than the modeled risk  $\text{VaR}_{0.99}(g_X(X))$ , as soon as the model distribution underestimates the tail risk probabilities for  $Z$ . Figure 2 establishes the same effect for exponentially distributed risks, thus showing that the issue persists for light-tailed risks. Taking  $\hat{\theta} = 1$ , we observe specifically that  $\text{VaR}_{0.99}(g_X(X)) = 0.7720$  but  $\text{VaR}_{0.99}(g_X(Z)) = 4.6098$  if  $Z \sim \text{Exp}(0.999)$ . Note that a 0.1% estimation error in the parameter  $\hat{\theta}$  leads here to an increase of almost 500% for the assessed risk. As a matter of fact, Figure 2 also shows that any benefits from optimizing VaR disappear as soon as  $\theta < \hat{\theta}$ , because then  $\text{VaR}_{0.99}(g_X(Z))$  becomes equal to the risk of the unoptimized position,  $\text{VaR}_{0.99}(Z)$ . In sharp contrast,  $\text{ES}_{0.975}(g_X(Z))$  ranges within the narrow interval  $[1.14619, 1.15216]$  for  $\theta \in [0.1, 1.5]$ . That is, the true risk  $\text{ES}_{0.975}(g_X(Z))$  deviates from the model value  $\text{ES}_{0.975}(g_X(X))$  by a mere 0.5% as long as the true value  $\theta$  is within  $\pm 50\%$  of the estimates value  $\hat{\theta}$ . Throughout that entire interval of  $\theta$ -values, the optimized position  $g_X(Z)$  provides a substantial and robust reduction of risk when compared

<sup>11</sup>The exponential( $\theta$ ) distribution function  $F$  with parameter  $\theta > 0$  is specified as  $F(x) = 1 - e^{-\theta x}$ ,  $x \geq 0$ .

with the ES of the non-optimized position,  $\text{ES}_{0.975}(X) = 4.6889$ .<sup>12</sup>

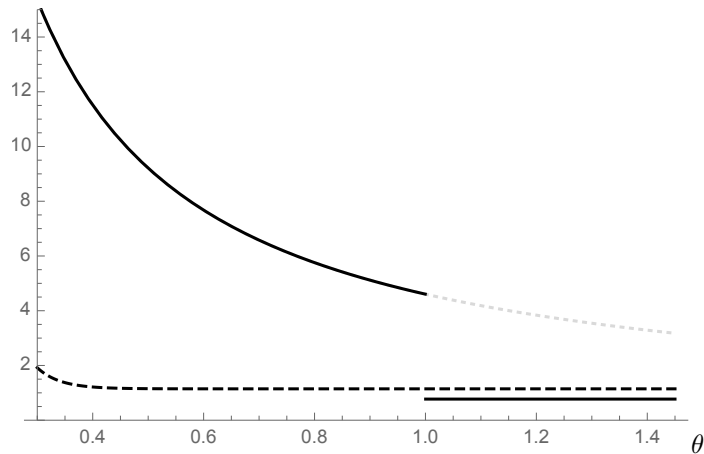


Figure 2: This plot shows  $\rho(g_X(Z))$  for  $\rho = \text{VaR}_{0.99}$  (solid) and  $\rho = \text{ES}_{0.975}$  (dashed), if  $Z$  is exponentially distributed with parameter  $\theta$ . We assume that  $X$  has an exponential distribution with parameter  $\hat{\theta} = 1$ . The dotted grey curve corresponds to the VaR of the unoptimized position,  $\text{VaR}_{0.99}(Z)$ , which coincides with  $\text{VaR}_{0.99}(g_X(Z))$  for  $\theta < \hat{\theta}$ .

The robustness of risk measures becomes important when risk measurement is combined with statistical estimation. This observation is at the core of the comparative discussion of the robustness of various risk measures; see e.g., [Cont et al. \(2010\)](#). In the following numerical experiment, we illustrate the impact of optimization on robustness. To this end, we refine the preceding setup by allowing for the statistical estimation of the parameter  $\theta$ . That is, we generate  $n$  iid realizations of  $Z$  and compute the maximum likelihood estimate  $\hat{\theta}$  from those realizations. Based on the estimated value of  $\hat{\theta}$ , we compute the optimizing function  $g_X$  and then compare the true risk value  $\rho(g_X(Z))$  to the perceived risk value  $\rho(g_X(X))$ . For each  $n$ , we repeat this procedure 10,000 times and compute the mean-squared error, i.e., the average of  $|\rho(g_X(Z)) - \rho(g_X(X))|^2$ , of all 10,000 sample points. As the number  $n$  of iid realizations of  $Z$  increases, the estimate  $\hat{\theta}$  becomes ever more accurate, and we may expect the mean-squared error of the risk differences to decrease; this is indeed true for the case of ES, but not true for the case of VaR, because VaR is not robust against optimization. [Figure 3](#) shows the corresponding mean-squared errors as a function of  $n$  for the case of Pareto-distributed risks. [Figure 4](#) shows the analogous computations for exponentially distributed risks. Both figures illustrate that ES massively outperforms VaR.

<sup>12</sup>Since  $\text{ES}_{0.975}(Z) \approx \text{VaR}_{0.99}(Z)$  for all  $\theta > 0$  as studied by [Li and Wang \(2019\)](#), we did not include the plot of  $\text{ES}_{0.975}(Z)$  in [Figure 2](#).

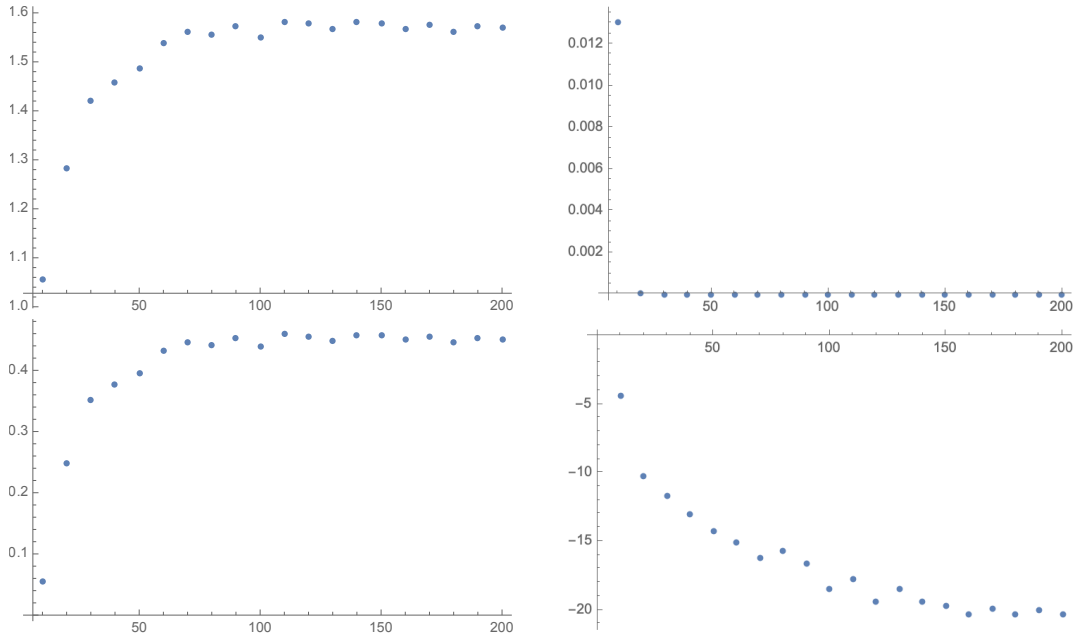


Figure 3: Mean-squared errors of 10,000 independent sample points of  $\rho(g_X(Z))$  and  $\rho(g_X(X))$ , each with a maximum likelihood estimator  $\hat{\theta}$  computed from  $n$  iid realizations of the Pareto(5)-distributed risk factor  $Z$ . The horizontal axis shows the number  $n$ . The case  $\rho = \text{VaR}_{0.99}$  can be found on the left,  $\rho = \text{ES}_{0.975}$  is on the right. The two top panels plot the mean-squared error (original values), the bottom ones correspond to their log-transforms.

## 8 Discussions and remarks

### 8.1 Implications of our results on regulatory risk measures

In both the banking and the insurance sectors, VaR and ES are competing regulatory risk measures for solvency capital calculation; see, for instance, [BCBS \(2016\)](#) from the Basel Committee on Banking Supervision and [IAIS \(2014\)](#) from the International Association of Insurance Supervisors. In this paper, with the new notion of robustness, we see that for the optimization problem (8), VaR is generally not robust whereas ES is. This provides strong support for the use of ES in optimization problems, in addition to its convexity which is very well recognized in the literature. These results further support the transition from VaR to ES made by [BCBS \(2016\)](#) from a novel theoretical perspective.

Our observations on the VaR vs. ES issue can be explained intuitively. From the proof of [Theorem 1](#), the VaR optimized positions always have a jump at the  $p$ -quantile level, and the optimized position can roughly be interpreted as a portfolio exhibiting a large loss with a small probability (e.g. selling a large volume of far out-of-the-money call options). This reflects the fact that “VaR does not capture the tail risk” as indicated already by many academics and regulators (see e.g. [Danielsson et al. \(2001\)](#), [Embrechts et al. \(2014\)](#), [Emmer et al. \(2015\)](#))

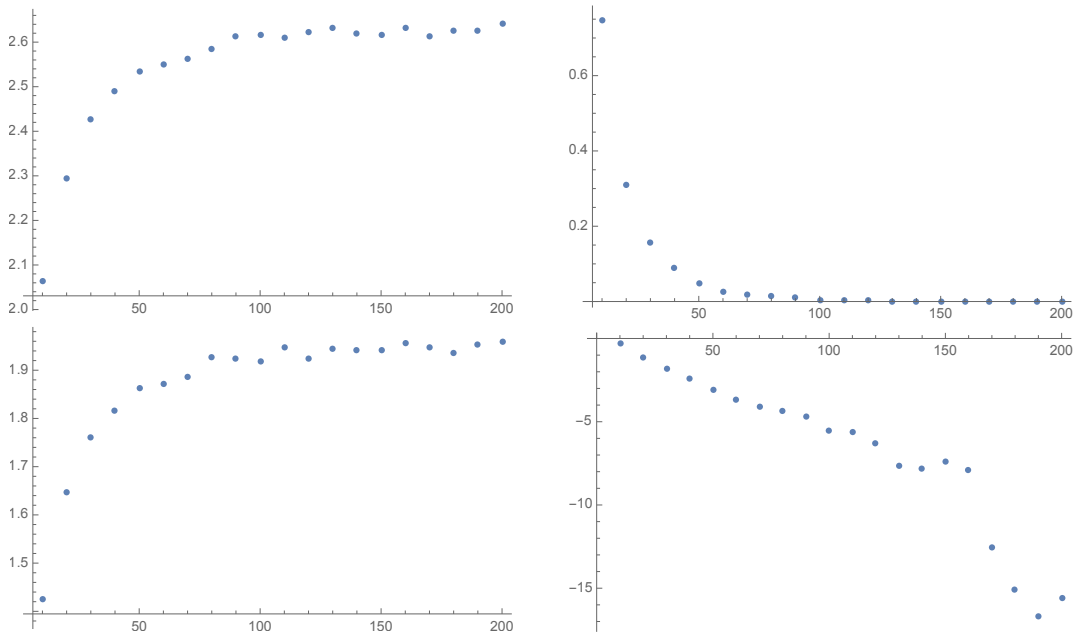


Figure 4: Mean-squared errors of 10,000 independent sample points of  $\rho(g_X(Z))$  and  $\rho(g_X(X))$ , each with a maximum-likelihood estimator  $\hat{\theta}$  computed from  $n$  iid realizations of the Exp(1)-distributed risk factor  $Z$ . The horizontal axis shows the number  $n$ . The case  $\rho = \text{VaR}_{0.99}$  can be found on the left,  $\rho = \text{ES}_{0.975}$  is on the right. The two top panels plot the mean-squared error (original values), the bottom ones correspond to their log-transforms.

and BCBS (2016)). If there is model uncertainty around this  $p$ -quantile level, even if small, it ruins completely the optimality of the position. This can explain the failure of the investment strategies (based on beliefs in small probabilities of default) of many of the larger banks before the 2008 financial crisis; see, for instance, the report by Acharya et. al. (2010) on this matter. We note that the optimized positions for VaR and ES may have similar forms (Proposition 3 and 4). By definition, however, ES does not ignore the values of the tail part of the optimal allocation (in contrast to VaR), and this explains why the corresponding value of the risk measure is not underestimated.

There are extensive discussions on the robustness of VaR and ES (although not in the context of optimization of this paper), and it may be fair to keep a balanced view. One important issue on the robustness of ES is the difficulty arising from perturbations which yield probability distributions that may have infinite first moment; this is why in Corollary 1, the robustness of ES with respect to  $\pi_n^q$  requires a condition on the growth rate of  $v$  and  $w$ . Therefore, in the minimization of ES from historical data, one needs to always make suitable integrability assumptions, or otherwise minimizing ES may be as problematic as the case of VaR. Infinite mean models are not of a purely academic nature in risk management; see for instance Nešlehová et al. (2006) in the context of operational risk and Weitzman (2009) related to the economics of

climate change. For recent academic discussions on various issues related to the desirability of VaR and ES in banking and insurance regulation, we refer to [Kou and Peng \(2016\)](#), [Fissler and Ziegel \(2016\)](#), [Embrechts et al. \(2018\)](#), [Armstrong and Brigo \(2018\)](#) and the references therein.

In the simulation examples of Section 7, we considered point estimation for a risk parameter of interest. As an alternative, we may adapt a Bayesian framework for the model parameter; see e.g., [Cakmak et al. \(2020\)](#) for recent development on Bayesian optimization for risk measures. One may wonder whether the lack of robustness against optimization for VaR is resolved by using a Bayesian approach to address statistical uncertainty. Since the Bayesian optimization problem rarely has analytical solutions, it is unclear whether a similar observation as in Theorem 1 would still hold. We leave robustness against optimization in a Bayesian framework for future research.

## 8.2 Remarks on the formulation of robustness

In this section we supply some further remarks on the relations between our definition of robustness and related notions in the literature on optimization and model uncertainty.

We start by considering the problem of solvency capital calculation of a firm as set forth in the Basel III/IV and Solvency II agreements. Suppose that  $Z$  is the true (but unknown) model and  $g_Z \in \mathcal{G}_Z(\rho)$ ; see (2) with  $X = Z$ . In solvency capital calculation, the following quantities have different physical meanings:

- (a)  $\rho(g_X(X))$ : the *perceived* risk value (solvency capital requirement) optimized for  $X$  by the firm;
- (b)  $\rho(g_Z(Z))$ : the *idealistic* risk value optimized for  $Z$  as if  $Z$  were known;
- (c)  $\rho(g_X(Z))$ : the *actual* risk value of the model  $Z$ , but the optimization is made for  $X$ .

Among the above quantities, the idealistic risk value  $\rho(g_Z(Z))$  represents what would be the best-case if the true model were known. Since the true model is not known, this value is not available and hence irrelevant for the solvency capital calculation. Therefore, for solvency risk management purposes, we are interested in the *solvency gap*

$$\underbrace{\rho(g_X(Z))}_{\text{actual risk}} - \underbrace{\rho(g_X(X))}_{\text{perceived risk}}, \quad (16)$$

not the *optimality gap*

$$\underbrace{\rho(g_Z(Z))}_{\text{idealistic optimum}} - \underbrace{\rho(g_X(Z))}_{\text{actual risk}}, \quad (17)$$

nor the *optimality shift*

$$\underbrace{\rho(g_Z(Z))}_{\text{idealistic optimum}} - \underbrace{\rho(g_X(X))}_{\text{perceived optimum}}. \quad (18)$$

Note that both (17) and (18) involve  $\rho(g_Z(Z))$  which is not relevant for solvency considerations. In the optimization literature, the continuity of the set mapping  $Z \mapsto \mathcal{G}_Z(\rho)$ , as well as that of the function  $Z \mapsto \rho(g_Z(Z))$ , is referred to as the problem of stability, i.e. how do the optimal solutions and the optimality shift (18) change when the underlying model changes from  $X$  to  $Z$ ; see, e.g., [Bonnans and Shapiro \(2000\)](#) and the references therein.

Let us further illustrate our notion of robustness by means of the following two examples.

**Example 6.** Suppose that the model  $X$  leads to the unique optimal decision  $g_X(x) = x_0 \in \mathbb{R}$ , which means fully liquidating this asset or a perfect hedge. In this case,  $g_X$  is a constant function, and hence  $Z \mapsto \rho(g_X(Z))$  is a constant mapping, thus always robust according to our definition. In other words, the solvency gap (16) will be zero, no matter what the optimizer for the true risk  $Z$  is. Hence, model uncertainty is irrelevant for the calculation of the solvency capital. On the other hand, if the true model  $Z$  is not equal to  $X$  and liquidating the asset is not optimal for  $Z$ , then we have  $\rho(g_Z(Z)) < \rho(x_0)$ ; thus the optimality shift (18) will be strictly negative. Therefore, the solvency gap in (16) is the right notion to look at in this scenario, not the optimality shift.

**Example 7.** Suppose that  $X$  and  $Z$  are similar in the sense that  $g_X(X)$  and  $g_Z(Z)$  are identically distributed, but  $g_X(Z)$  and  $g_X(X)$  are not identically distributed. In this case, we have  $\rho(g_X(X)) = \rho(g_Z(Z))$  for any law-invariant risk measure  $\rho$ , such as  $\text{VaR}_p$  or  $\text{ES}_p$ , but  $\rho(g_X(Z)) > \rho(g_Z(Z)) = \rho(g_X(X))$  since  $g_X$  is generally not optimal for  $Z$ . Clearly, the solvency gap (16) is strictly positive and the optimality shift (18) vanishes. In this situation, the regulator is concerned about the insolvency issue regarding model uncertainty. Indeed, the true risk value  $\rho(g_X(Z))$  is larger than the perceived risk value  $\rho(g_X(X))$ , which means that the solvency capital is insufficient. Note however that there is no gap between  $\rho(g_Z(Z))$  and  $\rho(g_X(X))$ . Therefore, also here, the solvency gap is the right notion to study rather than the optimality shift.

*Remark 3.* In the following few remarks, we provide various other interesting comments on our formulation of robustness against optimization.<sup>13</sup>

1. **Triangular inequality.** Note the triangular inequality

$$|\rho(g_X(Z)) - \rho(g_X(X))| \leq |\rho(g_X(Z)) - \rho(g_Z(Z))| + |\rho(g_Z(Z)) - \rho(g_X(X))|.$$

---

<sup>13</sup>We thank an anonymous referee for discussions on these issues, and in particular, pointing out the relationship of our notion and the classic ones via the triangular inequality.



This inequality suggests that if both the optimality shift in (18) and the optimality gap in (17) converges to 0 as  $Z \rightarrow X$  in  $\pi$ , then we have the solvency gap converges to 0, and thus robustness against optimization. However, the opposite does not hold, as illustrated by the situation in Example 6, where robustness is always guaranteed, since the optimizer for  $X$  is to fully liquidate the asset, although the optimality shift is non-zero. Hence, the study of robustness against optimization is not equivalent to the combined study of the optimality shift and stability. In other words, continuity in both optimality gap and optimality shift is sufficient but not necessary for robustness against optimization.

2. **Alternative ways to formulate robustness.** There are some alternative ways to formulate the notion of robustness in Definition 2. We discuss them and explain the advantages of our formulation.

- (a) One may use uncertainty on the set of probability distributions instead of that on the set of random vectors. There are a few advantages to consider the misspecification of the random vector rather than its distribution. First, our framework is general in the sense that there is no restriction to law-invariant risk measures or utility functions. For the notion of robustness studied in this paper, a probability measure does not need to be specified (non-law-invariant risk measures include e.g., the margin requirement risk measure used by the Chicago Mercantile Exchange; see McNeil et al. (2015, Section 2.3)). Second, our framework is flexible as we can easily incorporate the misspecification of distribution by using a metric on the set of distribution (which is a pseudo-metric on the set of random variables). Third, in the proof of several results in the paper on the robustness and non-robustness of risk measures in Sections 4 and 5, we need to obtain equalities in the almost sure sense ( $\omega$ -wise equalities) to locate the unique form of  $g$ .
- (b) An alternative to Definition 2 would be to require all, instead of one, optimizing functions  $g_X$  to satisfy  $\pi$ -continuity of  $Z \mapsto \rho(g_X(Z))$  at  $Z = X$ . This requirement would be stronger than the current Definition 2. For our main result, Theorem 1, stating VaR is generally *not robust*, the current formulation in Definition 2 gives a stronger result. Moreover, requiring continuity of all optimizing functions may lead to pathological statements. For instance, suppose that  $g_X \in \mathcal{G}$  is a continuous optimizing function for ES (or any other risk measure) and  $X \in L^1$ . If one modifies  $g_X$  on a set of  $\mu_X$ -measure zero (such as the set of rational numbers), then the resulting function is still optimal, but robustness fails, and rightly so.

3. **Limitations.** Robustness against optimization that we study in this paper is a desirable

notion of robustness, but it should be seen as a necessary, but generally not sufficient, condition for being a good risk measure to use in the context of optimization. For instance, even with the continuity of  $g_X(Z)$  at  $Z = X$ , small-size perturbations in the model may lead to enormous changes in the risk assessment in a practical example, as our notion does not quantify sensitivity of the risk value, which will be a future research direction.

4. **Optimization over stochastic processes.** One can consider continuous-time models where optimizers are chosen over a set of stochastic processes (e.g., admissible trading strategies). Our framework and discussions can be extended to such problems, as long as the optimizers are functions of the random source  $X$ , be it finite dimensional or infinite dimensional. In fact, in many classic financial models, the continuous-time optimization problem (such as the hedging example above) can be translated into a single-period optimization problem via the martingale approach as we see Example 4.

## 9 A connection to distributionally robust optimization

We conclude this paper by discussing our notion of robustness in the context of *distributionally robust optimization*. Our results in Section 4 show that VaR is generally not robust for (8). In a classic setting of distributionally robust optimization (e.g. Quaranta and Zaffaroni (2008), Zhu and Fukushima (2009), Blanchet and Murthy (2019)), the objective functional itself is evaluated under the worst-case value over a set of possible models representing uncertainty. By taking the worst-case value of the objective, model uncertainty is incorporated into the optimization problem. Some other relevant results on VaR and ES with the worst-case approach can be found in Hu and Hong (2013) and Zymler et al. (2013). We wonder whether robustness against optimization of risk measures would be improved by taking such an approach.

To formulate this consideration mathematically, let  $\rho$  be a compatible objective functional for an uncertainty triplet  $(\mathcal{G}, \mathcal{Z}, \pi)$  and  $X \in \mathcal{Z}$ . We look at the following optimization problem, which is a robust version of (1),

$$\text{to minimize: } \sup_{\pi(Y, X) \leq \varepsilon} \rho(g(Y)) \quad \text{subject to } g \in \mathcal{G}, \quad (19)$$

where  $\varepsilon > 0$ . Denote by  $\mathcal{G}_X(\rho, \varepsilon)$  the set of functions  $g \in \mathcal{G}$  minimizing (19). Clearly, if we allow  $\varepsilon = 0$  in (19), then  $\mathcal{G}_X(\rho, 0) = \mathcal{G}_X(\rho)$  and we are back in the setting of Section 2. In the problem (19), an investor is interested in the risk measure value  $\rho(g(Z))$  of the risky position  $g(Z)$ , in which  $Z$  is the unknowable true model. Therefore, similarly to Definition 2, we say that the objective functional  $\rho$  is robust against optimization for the setting (19) if there exists

$g_X \in \mathcal{G}_X(\rho, \varepsilon)$  such that the function  $Z \mapsto \rho(g_X(Z))$  is  $\pi$ -continuous at  $Z = X$ .

Unfortunately, the minimax problem (19) is not easy to solve analytically, even for the representative settings in Section 3 and in the cases of VaR and convex risk measures. Typically, a linear programming approach has to be applied for such problems. As convex risk measures are already shown to be generally robust in Section 5, its distributionally robust version is also generally robust; we thus focus on the question of whether VaR becomes more robust in this context. Our results in this section should be understood as exploratory rather than conclusive.

To obtain analytic results, we look at a simple one-dimensional case of (8), by letting

$$\mathcal{G} = \{g \in \mathcal{G}_1 : 0 \leq g \leq m, \mathbb{E}[\gamma(X)g(X)] \geq x_0\}, \quad (20)$$

where  $x_0$  and  $m$  are two constants satisfying  $0 \leq x_0 < m\mathbb{E}[\gamma(X)]$ . We choose  $(\mathcal{Z}, \pi) = (L^\infty, \pi_1^\infty)$  and formulate the optimization problem

$$\text{to minimize: } \sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(g(Y)) \text{ over } g \in \mathcal{G}. \quad (21)$$

Similarly to Section 4, denote by  $q$  the minimum of (21), that is,

$$q = \inf \left\{ \sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(g(Y)) : 0 \leq g \leq m, \mathbb{E}[\gamma(X)g(X)] \geq x_0 \right\}.$$

We make the following stronger assumption.

**Assumption D.**  $q > 0$ ,  $1/2 \leq p < 1$ ,  $X$  has a decreasing density on its support and  $\gamma$  is an increasing function of  $X$ .<sup>14</sup>

Fortunately, with Assumption D, we are able to obtain an explicit form of the solution to Problem (21), allowing us to compare the corresponding robustness property with the results we obtained in Section 4.

**Proposition 5.** *For  $\mathcal{G}$  in (20), under Assumption D, Problem (21) admits a solution of the form*

$$g_X(x) = m\mathbb{1}_{\{x > c + \varepsilon\}} + q\mathbb{1}_{\{x \leq c + \varepsilon\}}, \quad x \in \mathbb{R}, \quad \text{where } c = \text{VaR}_p(X). \quad (22)$$

With the solution  $g_X$  in Proposition 5, the continuity of VaR mentioned in Section 3 implies that the mapping  $Z \mapsto \text{VaR}_p(g_X(Z))$  is  $\pi_1^\infty$ -continuous at  $Z = X$ . As a consequence,  $\text{VaR}_p$  is robust against optimization for the setting (19). This observation is in sharp contrast

---

<sup>14</sup>The monotonicity of  $\gamma$  as a function of  $X$  has a simple economic meaning. Recall that  $X$  represents the loss of an asset. Hence, Assumption D requires that the pricing density is larger when the asset has a larger loss. This requirement is satisfied by classic equilibrium models in the notion of Arrow-Debreu (Arrow and Debreu (1954)).

with Theorem 1, where we see that  $\text{VaR}_p$  is not robust for  $(\mathcal{G}, L^\infty, \pi_1^\infty)$  under some very weak assumptions (which does not conflict Assumption D). Therefore, at least for the special setting (19), the modified optimization problem (21) improves the robustness of VaR. It is unclear how this result can be generalized to other optimization problems, as analytic results for (19) are rarely available.

Although  $\text{VaR}_p$  becomes robust in the setting (21), its optimizing function takes a similar form as in Proposition 3, that is, a distribution with a jump and a big loss with small probability. Since the distribution of  $g_X(X)$  in (22) has a jump at its  $(p + \varepsilon)$ -quantile, this type of optimizing functions is highly undesirable and is subject to considerable model uncertainty if  $\varepsilon$  is small; see the discussions in Section 8.

**Acknowledgements.** The authors thank the Area Editor Paul Glasserman, an Associate Editor, two anonymous referees, Rama Cont, Xiaoxue Deng, Tobias Fissler, Liyuan Lin, Stéphane Loisel, Marcel Nutz, and Philip Protter for insightful comments on an early version of the paper. In particular, the term “robustness against optimization” was suggested by Paul Glasserman. RW acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2018-03823, RGPAS-2018-522590) and from the Center of Actuarial Excellence Research Grant from the Society of Actuaries.

## References

- Acharya, V. V., Cooley, T. and Richardson, M. (2010). *Manufacturing Tail Risk: A Perspective on the Financial Crisis of 2007-2009*. Now Publishers Inc.
- Armstrong, J. and Brigo, D. (2018). Rogue traders versus value-at-risk and expected shortfall. *Risk Magazine*, April 2018.
- Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, **53**(5), 941–973.
- Arrow, K. J. and Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, **22**(3), 265–290.
- Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.
- BCBS (2016). *Standards. Minimum Capital Requirements for Market Risk. January 2016*. Basel Committee on Banking Supervision. Basel: Bank for International Settlements.
- Cheridito, P. and Li, T. (2009). Risk measures on Orlicz hearts. *Mathematical Finance*, **19**(2), 189–214.
- Ben-Tal, A., El Ghaoui, L. and Nemirovski, A. (2009). *Robust Optimization*. Princeton University Press, New Jersey.

- Ben-Tal, A. and Teboulle, M. (1987). Penalty functions and duality in stochastic programming via  $\phi$ -divergence functionals. *Mathematics of Operations Research*, **12**, 224–240.
- Ben-Tal, A. and Teboulle, M. (2007). An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, **17**(3), 449–476.
- Bernard, C., He, X., Yan, J. A. and Zhou, X. Y. (2015). Optimal insurance design under rank-dependent expected utility. *Mathematical Finance*, **25**, 154–186.
- Blanchet, J. and Murthy, K. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, **44**(2), 565–600.
- Bonnans, J. F. and Shapiro, A. (2000). *Perturbation Analysis of Optimization Problems*. Springer Series in Operations Research. Springer-Verlag, New York.
- Cakmak, S., Astudillo, R., Frazier, P. and Zhou, E. (2020). Bayesian optimization of risk measures. *arXiv*: 2007.05554.
- Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, **10**(6), 593–606.
- Daniélsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking and Finance*, **26**(7), 1273–1296.
- Daniélsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O. and Shin, H. S. (2001). An academic response to Basel II. *LSE Special Paper Series May 2001*.
- Dunford, N. and Schwartz, J. (1958). *Linear Operators. Part I: General Theory*. Interscience Publishers, New York.
- Ekeland, I. and Temam, R. (1976). Convex analysis and variational problems. Translated from the French. *Studies in Mathematics and its Applications*, Vol. 1. North-Holland Publishing Co., Amsterdam-Oxford.
- Embrechts, P., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, **66**(4), 936–949.
- Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25–48.
- Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, **19**(4), 763–790.
- Embrechts, P. and Wang, R. (2015). Seven proofs for the subadditivity of Expected Shortfall. *Dependence Modeling*, **3**, 126–140.
- Emmer, S., Kratz, M. and Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, **18**(2), 31–60.
- Fissler, T. and Ziegel, J. F. (2016). Higher order elicibility and Osband’s principle. *Annals of Statistics*, **44**(4), 1680–1707.
- Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, **6**(4), 429–447.

- Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time*. Fourth edition, de Gruyter.
- Föllmer, H., Schied, A. and Weber, S. (2009). Robust preferences and robust portfolio choice. In: *Mathematical Modelling and Numerical Methods in Finance*, (eds.: P. Ciarlet, A. Bensoussan, Q. Zhang), Handbook of Numerical Analysis 15, 29–88.
- Goh, J. and Sim, M. (2010). Distributionally robust optimization and its tractable approximations. *Operations Research*, **58**(4), 902–917.
- Goodhart, C. A. (1984). Problems of monetary management: the UK experience. In *Monetary Theory and Practice* (pp. 91–121). Palgrave, London.
- He, X. D. and Zhou, X. Y. (2011). Portfolio choice via quantiles. *Mathematical Finance*, **21**(2), 203–231.
- Hu, Z. and Hong, L. J. (2013). Kullback-Leibler divergence constrained distributionally robust optimization. Available at Optimization Online.
- IAIS (2014). *Consultation Document December 2014. Risk-based global insurance capital standard*. International Association of Insurance Supervisors.
- Jayne, J. E. and Rogers, C. A. (1985). Borel selectors for upper semi-continuous set-valued maps. *Acta Mathematica*, **155**, 41–79.
- Kou, S. and Peng, X. (2016). On the measurement of economic tail risk. *Operations Research*, **64**(5), 1056–1072.
- Kou, S., Peng, X. and Heyde, C. C. (2013). External risk measures and Basel accords. *Mathematics of Operations Research*, **38**(3), 393–417.
- Krätschmer, V., Schied, A. and Zähle, H. (2014). Comparative and quantitative robustness for law-invariant risk measures. *Finance and Stochastics*, **18**(2), 271–295.
- Krätschmer, V., Schied, A. and Zähle, H. (2017). Domains of weak continuity of statistical functionals with a view toward robust statistics. *Journal of Multivariate Analysis*, **158**, 1–19.
- Li, H. and Wang, R. (2019). PELVE: Probability equivalent level of VaR and ES. *SSRN*: 3489566.
- McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, NJ: Princeton University Press.
- Natarajan, K., Pachamanova, D. and Sim, M. (2008). Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, **54**(3), 573–585.
- Nešlehová, J., Embrechts, P., and Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for operational risk. *Journal of Operational Risk*, 1(1), 3–25.
- Newey, W. K. and Powell, J. L. (1987) Asymmetric least squares estimation and testing. *Econometrica*, **55**(4), 819–847.
- Pflug, G. C. and Römisch, W. (2007). *Modeling, Measuring and Managing Risk*. World Scientific Publishing, Hackensack, NJ.
- Quaranta, A. G. and Zaffaroni, A. (2008). Robust optimization of conditional value at risk and portfolio selection. *Journal of Banking and Finance*, **32**(10), 2046–2056.

- Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.
- Rudin, W. (1976). *Principles of Mathematical Analysis*. Mac Graw-Hill.
- Sekine, J. (2004). Dynamic minimization of worst conditional expectation of shortfall. *Mathematical Finance*, **14**(4), 605–618.
- Shorack, G. (2000). *Probability for Statisticians*. Springer Texts in Statistics. Springer-Verlag, New York.
- Strathern, M. (1997). ‘Improving ratings’: audit in the British University system. *European Review*, **5**(3), 305–321.
- Wang, R. and Zitikis, R. (2020). An axiomatic foundation for the Expected Shortfall. *Management Science*, published online.
- Weitzman M. L. (2009). On modeling and interpreting the economics of catastrophic climate change. *The Review of Economics and Statistics*, 91(1), 1-19.
- Wiesemann, W., Kuhn, D. and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, **62**(6), 1358–1376.
- Zhu, S. and Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, **57**(5), 1155-1168.
- Zymler, S., Kuhn, D. and Rustem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming Series A*, **137**, 167–198.

## Planned E-Companion: Appendices

### A Proofs of theorems and propositions

#### A.1 Proofs in Section 2

*Proof of Proposition 1.* It suffices to show that the function  $Z \mapsto \rho(g(Z))$  is  $\pi$ -continuous. By definition, for any  $X, Y \in \mathcal{Z}$ ,  $\pi_{g(\mathcal{Z})}(g(X), g(Y)) = \pi(X, Y)$ . Thus the  $\pi_{g(\mathcal{Z})}$ -continuity of  $\rho$  is equivalent to the  $\pi$ -continuity of the function  $Z \mapsto \rho(g(Z))$ .  $\square$

*Proof of Proposition 2.* (i) It suffices to show that, as  $k \rightarrow \infty$ ,  $X_k \rightarrow X$  in  $\pi_n^\infty$  implies that  $g(X_k) \rightarrow g(X)$  in  $\pi_1^\infty$ . This is a direct consequence of the Heine-Cantor Theorem (see Theorem 4.19 of Rudin (1976)).

(ii)  $X_k \rightarrow X$  w.r.t.  $\pi_n^q$  implies that  $\{|X_k|^q\}_{k \in \mathbb{N}}$  is uniformly integrable and that  $X_k \rightarrow X$  in probability. It follows from the Continuous Mapping Theorem that  $g(X_k) \rightarrow g(X)$  in probability. Moreover, for sufficiently large  $c$ ,

$$\mathbb{E}\left[|g(X_k)|^q \mathbf{1}_{\{|g(X_k)| > c\}}\right] \leq C^q \mathbb{E}\left[|X_k|^q \mathbf{1}_{\{|X_k| > c/C\}}\right].$$

Therefore,  $(|g(X_k)|^q)$  is uniformly integrable and, in turn,  $g(X_k) \rightarrow g(X)$  w.r.t.  $\pi_1^q$ .

(iii) It suffices to show that, as  $k \rightarrow \infty$ ,  $X_k \rightarrow X$  in  $\pi_n^W$  implies that  $g(X_k) \rightarrow g(X)$  in  $\pi_1^W$ . This is a direct consequence of the Continuous Mapping Theorem.  $\square$

#### A.2 Proofs in Section 4

Since a rescaling of  $\gamma$  does not change the optimization problem (8), we will safely assume  $\mathbb{E}[\gamma(X)] = 1$  in the proofs of all results in Sections 4-6.

*Proof of Theorem 1.* In what follows, equalities and inequalities on functions on  $\mathbb{R}^n$  are understood as almost surely with respect to  $\mu_X$ , and essential suprema and expectations of these functions are taken under  $\mu_X$  (we use  $\mathbb{E}_X$  to emphasize the expectation with respect to  $\mu_X$ ).

There is nothing to show if the set  $\mathcal{G}_X(\rho)$  of minimizers is empty. Suppose that  $g_X \in \mathcal{G}$  is a minimizer to the problem (1). We will show that  $Z \mapsto \rho(g_X(Z))$  cannot be continuous at  $X$ , which gives the statement in the theorem.

We first show that the budget constraint is always binding, that is,

$$\mathbb{E}_X[\gamma g'_X] = x_0 \text{ for any optimizer } g'_X \text{ to (1)}. \tag{A.1}$$



Suppose  $\mathbb{E}_X[\gamma g'_X] > x_0$  for contradiction. Denote by  $\varepsilon = \mathbb{E}_X[\gamma g'_X] - x_0$ ,  $v_0 = \text{ess-sup}(v)$ , and let  $g''_X = (g'_X - \varepsilon) \vee v$ . Since  $v \leq g''_X \leq w$  and  $\mathbb{E}_X[\gamma g''_X] \geq \mathbb{E}_X[\gamma g'_X] - \varepsilon = x_0$ , we have  $g''_X \in \mathcal{G}$ . Moreover, since  $v_0 < \text{VaR}_p(g'_X(X))$  by Assumption **V**, we have

$$\begin{aligned} \text{VaR}_p(g''_X(X)) &\leq \text{VaR}_p((g'_X(X) - \varepsilon) \vee v_0) \\ &= \text{VaR}_p(g'_X(X) - \varepsilon) \vee v_0 \\ &= (\text{VaR}_p(g'_X(X)) - \varepsilon) \vee v_0 < \text{VaR}_p(g'_X(X)). \end{aligned}$$

This contradicts the optimality of  $g'_X$ . Hence, (A.1) holds.

Consider the probability space  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_X)$ , where  $\mathcal{B}(\mathbb{R}^n)$  is the Borel  $\sigma$ -field. Following Wang and Zitikis (2020), a set  $A \in \mathcal{B}(\mathbb{R}^n)$  is called a  $p$ -tail event for a measurable function  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $p \in (0, 1)$ , if  $\mu_X(A) = 1 - p$  and  $h(x) \geq h(x')$  for  $\mu_X$ -a.e.  $x \in A$  and  $x' \in A^c$ . The existence of the tail event in any atomless probability space is implied by Lemma A.3 of Wang and Zitikis (2020), noting that Assumption **G** guarantees that  $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_X)$  is an atomless probability space.

Let  $A$  be a  $p$ -tail event of  $g_X$  so that  $\mathbb{P}(X \in A) = \mu_X(A) = 1 - p$ . By the definition of  $\text{VaR}_p$ , we have

$$\text{VaR}_p(g_X(X)) = \text{ess-sup}(g_X|A^c), \quad (\text{A.2})$$

where  $\text{ess-sup}(g_X|A^c)$  is the essential supremum (with respect to  $\mu_X$ ) of  $g_X$  conditional on  $A^c$ . Define the function  $\hat{g} = g_X \mathbb{1}_{A^c} + w \mathbb{1}_A$ . It is clear that  $\hat{g} \geq g_X$ . Moreover, since  $\hat{g}$  and  $g_X$  only differ on the tail event  $A$ , (A.2) implies that  $\text{VaR}_p(\hat{g}(X)) = \text{VaR}_p(g_X(X))$ . Hence,  $\hat{g}$  is also a minimizer to (1). Note that if  $\hat{g} \neq g_X$ , then we have  $\mathbb{E}_X[\gamma \hat{g}] > \mathbb{E}_X[\gamma g_X] = x_0$ . By (A.1), the budget constraint is always binding and we conclude that  $\hat{g} = g_X$ . Thus,  $g_X \mathbb{1}_A = w \mathbb{1}_A$ .

Next, suppose by way of contradiction that the quantile function of  $g_X(X)$  is continuous at  $p$ . If a  $p$ -tail event  $A'$  of  $w$  is  $\mu_X$ -a.s. equal to  $A$ , then we have, using  $g_X \mathbb{1}_A = w \mathbb{1}_A = w \mathbb{1}_{A'}$ ,

$$\text{VaR}_p(w(X)) \leq \lim_{q \downarrow p} \text{VaR}_q(w(X)) = \lim_{q \downarrow p} \text{VaR}_q(g_X(X)) = \text{VaR}_p(g_X(X)),$$

a contradiction to  $\text{VaR}_p(g_X(X)) < \text{VaR}_p(w(X))$  in Assumption **V**. Hence, any  $p$ -tail event  $A'$  of  $w$  satisfies  $A' \neq A$ . Since both sets have the same probability, the set  $C := A' \setminus A$  must be such that  $\alpha := \mu_X(C) \in (0, 1 - p]$ .

Write  $a = \text{VaR}_p(g_X(X))$  and  $b = \text{VaR}_p(w(X))$ . For each  $\delta \in (0, \alpha)$ , let  $C_\delta$  be a subset of  $C$  such that  $\mu_X(C_\delta) = \delta$ ,  $A_\delta$  be a  $(p + \delta)$ -tail event of  $g_X$ , and  $B_\delta := A \setminus A_\delta$ . Note that  $w \geq b$  and  $g_X \leq a$  on  $C_\delta$  since  $C \subset A' \setminus A$ . Hence,  $w - g_X \geq b - a > 0$  on  $C_\delta$ . Moreover,  $C_\delta \cap B_\delta = \emptyset$ . Since  $\gamma > 0$ ,  $\mu_X(\gamma \geq \varepsilon) \rightarrow 1$  as  $\varepsilon \downarrow 0$ . As a consequence, we can freely choose  $C_\delta$  such that  $\gamma$  is

bounded away from 0 on  $C_\delta$  for each  $\delta \in (0, \alpha)$ . Since  $\gamma$  is bounded from above, we can let  $\ell, u$  be two constants such that  $0 < \ell < \gamma$  on  $C_{\alpha/2}$  and  $\gamma < u < \infty$  on  $B_{\alpha/2}$ .

Let  $g_\delta = a\mathbf{1}_{B_\delta} + w\mathbf{1}_{C_\delta} + g_X(1 - \mathbf{1}_{B_\delta \cup C_\delta})$ . In other words,  $g_\delta$  is obtained by decreasing the value of  $g_X$  to  $a$  on the set  $B_\delta$  of probability  $\delta$ , and increasing its value to  $w$  on the set  $C_\delta$  also of probability  $\delta$ . Clearly,  $v \leq g_\delta \leq w$ . Note that  $g_X \leq a$  on  $A^c$ , which implies  $g_\delta \leq a$  on  $A^c \setminus C_\delta$ . Moreover,  $g_\delta \leq a$  on  $B_\delta$ . Therefore,

$$\mathbb{P}(g_\delta(X) \leq a) \geq \mu_X(B_\delta) + \mu_X(A^c \setminus C_\delta) = 1 - p,$$

which gives  $\text{VaR}_p(g_\delta(X)) \leq a$ . Since the quantile of  $g_X(X)$  is continuous at  $p$ , there exists  $\delta_0 \in (0, \alpha/2)$  such that  $|g_X - a| < (b - a)\ell/u$  on  $B_{\delta_0}$ . Putting the above observations together, we have

$$\begin{aligned} \mathbb{E}[\gamma g_{\delta_0}] - \mathbb{E}[\gamma g_X] &= \mathbb{E}[\gamma(g_{\delta_0} - g_X)\mathbf{1}_{C_{\delta_0}}] - \mathbb{E}[\gamma(g_X - g_{\delta_0})\mathbf{1}_{B_{\delta_0}}] \\ &\geq (b - a)\mathbb{E}[\gamma\mathbf{1}_{C_{\delta_0}}] - \frac{\ell}{u}(b - a)\mathbb{E}[\gamma\mathbf{1}_{B_{\delta_0}}] \\ &\geq (b - a)\ell\delta_0 - \frac{\ell}{u}(b - a)u\delta_0 > 0. \end{aligned}$$

The facts that  $\text{VaR}_p(g_{\delta_0}(X)) \leq a = \text{VaR}_p(g_X(X))$  and  $\mathbb{E}_X[\gamma g_{\delta_0}] > \mathbb{E}_X[\gamma g_X] = x_0$  further guarantees that  $g_{\delta_0}$  is an optimizer to (1). However, this contradicts the fact that any optimizer to (1) needs to satisfy (A.1). This contradiction shows the desired conclusion that the quantile function of  $g_X(X)$  has a jump at  $p$ .

For  $\varepsilon > 0$ , let  $\tilde{A}_\varepsilon = \{x \in \mathbb{R}^n : d(x, A) \leq \varepsilon\}$ , where  $d$  is the Euclidean distance. For each  $y \in \tilde{A}_\varepsilon$ , let  $f_\varepsilon(y)$  be a Borel function which maps  $y$  to one of its nearest point in  $A$ ; see e.g., [Jayne and Rogers \(1985\)](#) for the existence of the Borel selector. Define the random variables  $Z_\varepsilon$  by

$$Z_\varepsilon = f_\varepsilon(X)\mathbf{1}_{\{X \in \tilde{A}_\varepsilon\}} + X\mathbf{1}_{\{X \in \tilde{A}_\varepsilon^c\}}.$$

Note that  $\pi_n^\infty(Z_\varepsilon, X) \leq \varepsilon$ . Hence,  $Z_\varepsilon \rightarrow X$  as  $\varepsilon \downarrow 0$  in  $\pi_n^\infty$ , which is the strongest metric  $\pi$  that we consider. Moreover, by Assumption **G**,  $X$  has positive density over its support which is a convex set, which implies  $\mathbb{P}(Z_\varepsilon \in A) = \mathbb{P}(X \in \tilde{A}_\varepsilon) > \mathbb{P}(X \in A) = 1 - p$ . Also note that if  $Z_\varepsilon \in A$ , then  $g_X(Z) \geq \lim_{q \downarrow p} \text{VaR}_q(g_X(X))$ . Hence,

$$\text{VaR}_p(g_X(Z_\varepsilon)) \geq \lim_{q \downarrow p} \text{VaR}_q(g_X(X)) > \text{VaR}_p(g_X(X)),$$

showing that  $Z \mapsto \rho(g_X(Z))$  is not  $\pi$ -continuous at  $X$ . □

*Remark 4.* The assumption  $\text{ess-sup}(v) < \text{VaR}_p(X; \mathcal{G})$  in Assumption **V** is not essential. As we

see from the proof, this assumption is used to show two conditions. First, it is used to show (A.1); i.e., the budget constraint is binding. Second, it is used to guarantee that  $v \leq g_\delta \leq w$  for  $\delta$  small enough. These two conditions are both natural and quite weak. On the other hand, the assumption that  $\gamma$  is bounded above is only used to guarantee that  $\gamma < u$  for some  $u > 0$  on the set  $B_\delta$  with probability  $\delta \downarrow 0$ . Note that if  $g_X(X)$  is continuously distributed, then the set  $B_\delta$  is  $\mu_X$ -a.s. equal to  $\{x \in \mathbb{R}^n : \text{VaR}_p(g_X(X)) < g_X(x) < \text{VaR}_{p+\delta}(g_X(X))\}$ . Hence, in this case it suffices to assume that  $\gamma$  is bounded from above in any small neighbourhood of  $\{x \in \mathbb{R}^n : g_X(x) = \text{VaR}_p(g_X(X))\}$ . This assumption is practically always satisfied.

The following lemma is needed to show Proposition 3.

**Lemma A.1.** *Under Assumptions G and V, Problem (9) admits at least one solution.*

*Proof of Lemma A.1.* Define  $\mathbb{Q}$  through  $d\mathbb{Q}/d\mathbb{P} = \gamma$  and let  $\mu = \mathbb{Q} \circ X^{-1}$ . The set  $\mathcal{G}$  is then a uniformly integrable subset of  $L^1(\mu)$ . Let  $\{g_n\}_{n \in \mathbb{N}}$  be a minimizing sequence for  $\text{VaR}_p$  in  $\mathcal{G}$ . By the Dunford-Pettis and Eberlein-Šmulian theorems (Theorems IV.8.9 and V.6.1 of Dunford and Schwartz (1958)), there exists a subsequence  $\{g_{n_k}\}_{k \in \mathbb{N}}$  that converges weakly in  $L^1(\mu)$  to some function  $g_X \in L^1(\mu)$ . Since  $\mathcal{G}$  is convex and closed in  $L^1(\mu)$ , we get  $g_X \in \mathcal{G}$ . Moreover, weak convergence in  $L^1(\mu)$  implies clearly that the laws of  $g_{n_k}(X)$  converge weakly to the one of  $g_X(X)$ . But  $\text{VaR}_p$  is a left-hand quantile and hence lower semicontinuous with respect to weak convergence (see, e.g., Exercise A.6.1 in Föllmer and Schied (2016)). This proves that  $g_X$  is optimal.  $\square$

*Proof of Proposition 3.* It is straightforward to check that  $\text{VaR}_p(g_X(X)) \leq q$ . By Lemma A.1, Problem (9) has at least one optimizer. Let  $g \in \mathcal{G}$  be an optimizer to (9). Since  $\text{VaR}_p(g(X)) = q$  and  $g(X) \leq X$ , we have, in the sense of  $\mu_X$ -a.s.,

$$g(X) \leq X \mathbf{1}_A + (X \wedge q) \mathbf{1}_{A^c},$$

where  $A$  is a  $p$ -tail event of  $g(X)$ . Clearly, by taking an equality in the above inequality will only increase  $\mathbb{E}_X[\gamma g]$  while maintaining  $\text{VaR}_p(g(X)) \leq q$ , and it does not affect the optimality of  $g$ . Moreover, as we have seen in the proof of Theorem 1, the budget constraint is binding; this implies that we cannot strictly increase  $\mathbb{E}_X[\gamma g]$  while maintaining  $\text{VaR}_p(g(X)) \leq q$ . Therefore, it has to be  $g(X) = X \mathbf{1}_A + (X \wedge q) \mathbf{1}_{A^c}$ . Note that

$$\begin{aligned} \mathbb{E}[\gamma(X)g(X)] &= \mathbb{E}[\gamma(X)X \mathbf{1}_A] + \mathbb{E}[\gamma(X)(X \wedge q) \mathbf{1}_{A^c}] \\ &= \mathbb{E}[\gamma(X)X] - \mathbb{E}[\gamma(X)(X - q)_+ \mathbf{1}_{A^c}] = \mathbb{E}[\gamma(X)X] - \mathbb{E}[Y_+ \mathbf{1}_{A^c}]. \end{aligned} \quad (\text{A.3})$$

Maximizing the above term over  $A$  satisfying  $\mathbb{P}(A) = 1 - p$ , it is clear that the maximum of  $\mathbb{E}[\gamma(X)g(X)]$  is attained when  $Y_+$  takes its smallest values on  $A^c$ . In other words,  $A$  is a  $p$ -tail event of  $Y_+$ . Moreover, we have  $q < \text{VaR}_p(X)$  by Assumption **V**. Hence,  $Y = Y_+ > 0$  on  $A$ , and  $A$  is also a  $p$ -tail event of  $Y$ . Using again the fact that the budget constraint is binding, any function  $g'$  that does not maximize  $\mathbb{E}_X[\gamma g']$  with fixed  $\text{VaR}_p(g'(X)) = q$  cannot be an optimizer. Since the  $p$ -tail event of  $Y_+$  is unique by  $\mathbb{P}(Y \leq \text{VaR}_p(Y)) = p$ , we know that  $g = g_X$  in (10) is the unique  $g$  with  $\text{VaR}_p(g(X)) = q$  such that  $\mathbb{E}_X[\gamma g] = x_0$ . As a consequence,  $g = g_X$  is the  $\mu_X$ -a.s. unique solution to (9).

Finally, we show (11). Since  $A$  is a  $p$ -tail event of  $Y_+$ , we know that  $A^c$  is a  $(1 - p)$ -tail event of  $-Y_+$ . Further, we have  $\mathbb{E}[-Y_+|A^c] = \text{ES}_{1-p}(-Y_+)$  by Lemma A.7 of Wang and Zitikis (2020). Using the fact that the budget constraint is binding and (A.3), we obtain

$$\begin{aligned} x_0 &= \mathbb{E}[\gamma(X)g_X(X)] \\ &= \mathbb{E}[\gamma(X)X] - \mathbb{E}[Y_+ \mathbf{1}_{A^c}] \\ &= \mathbb{E}[\gamma(X)X] + p\mathbb{E}[-Y_+|A^c] = \mathbb{E}[\gamma(X)X] + p\text{ES}_{1-p}(-Y_+). \end{aligned}$$

This gives the desired equality (11). □

### A.3 Proofs in Section 5

*Proof of Theorem 2.* In a first step, we show the existence of a minimizer in  $\mathcal{G}$ . Clearly,  $\mathcal{G}$  is nonempty. Let  $\{g_n\}_{n \in \mathbb{N}}$  be a sequence in  $\mathcal{G}$  such that  $\rho(g_n(X))$  converges to  $\lambda := \inf_{g \in \mathcal{G}} \rho(g(X))$ . Since  $g_n$  takes values between  $v$  and  $w$ , a standard Komlós-type argument (e.g., Lemma 1.70 in Föllmer and Schied (2016)), allows us to pass to a sequence  $\{\tilde{g}_n\}_{n \in \mathbb{N}}$  of convex combinations of the  $g_n$  such that  $\tilde{g}_n$  converges  $\mu_X$ -a.s. to some function  $g_0$ . Dominated convergence yields that

$$\mathbb{E}[\gamma(X)g_0(X)] = \lim_{n \uparrow \infty} \mathbb{E}[\gamma(X)\tilde{g}_n(X)] \geq \liminf_{n \uparrow \infty} \mathbb{E}_X[\gamma(X)g_n(X)] \geq x_0. \quad (\text{A.4})$$

Hence,  $g_0$  belongs to  $\mathcal{G}$ . The convexity of  $\rho$  implies that  $\rho(\tilde{g}_n(X))$  converges to the infimum value  $\lambda$ . Moreover, since  $\rho$  enjoys the Fatou property due to (13), we have  $\rho(g_0(X)) \leq \liminf_n \rho(\tilde{g}_n(X))$ . Therefore  $g_X := g_0$  is a minimizer in  $\mathcal{G}$ .

Now we derive the structure of  $g_X$ . To this end, consider  $\ell(x) := \sup_{y \geq 0} (xy - \varphi(y))$ . Then  $\ell$  is convex, nondecreasing, nonconstant, and finite on  $\mathbb{R}$ , because  $\varphi$  has superlinear growth. Moreover, Theorem 4.2 in Ben-Tal and Teboulle (2007) (see also Lemma 1 in Ben-Tal and

Teboulle (1987)) provides the following dual representation,

$$\rho(Y) = \inf_{z \in \mathbb{R}} (\mathbb{E}[\ell(Y + z)] - z), \quad Y \in L^\infty. \quad (\text{A.5})$$

We claim that the infimum in (A.5) is actually attained. To see why, note first that our assumptions on  $\varphi$  imply that there exists  $y_0 > 1$  such that  $\varphi(y_0) < \infty$ , which implies that the slope of  $\ell$  is at least  $y_0$  for sufficiently large  $x$ . This yields an upper bound on the range of all  $z$  that contribute to the infimum in (A.5). Second, there is  $y_1 < 1$  with  $\varphi(y_1) < \infty$ , which implies that the slope of  $\ell$  is at most  $y_1$  for sufficiently large negative  $x$ . This gives a lower bound on  $z$ . The continuity of  $z \mapsto \mathbb{E}[\ell(Y + z)] - z$  now yields our claim.

Now let  $z^*$  be such that  $\rho(g_X(X)) = \mathbb{E}[\ell(g_X(X) + z^*)] - z^*$ . Then,

$$\inf_{g \in \mathcal{G}} \rho(g(X)) \leq \inf_{g \in \mathcal{G}} (\mathbb{E}[\ell(g(X) + z^*)] - z^*) \leq \mathbb{E}[\ell(g_X(X) + z^*)] - z^* = \inf_{g \in \mathcal{G}} \rho(g(X)).$$

Hence,  $g_X$  minimizes  $\mathbb{E}[\ell(g(X) + z^*)]$  over  $g \in \mathcal{G}$ . We are thus in the context of Theorem 4 (optimizing expected loss), whose proof yields the form of  $g_X$  as the minimizer of  $\mathbb{E}[\ell(g(X) + z^*)]$  over  $g \in \mathcal{G}$ . Now suppose that  $Z_n \in L_n^0$  are random variables whose laws converge to the one of  $X$ . By Skorokhod embedding, we may assume without loss of generality that  $Z_n \rightarrow X$  holds  $\mathbb{P}$ -a.s. The robustness of  $\rho$  now follows as in the proof of Theorem 4 by using the fact that  $\rho$  enjoys the so-called Lebesgue property, which in turn is a consequence of Exercises 4.2.3 and 4.3.4 in Föllmer and Schied (2016).  $\square$

*Proof of Corollary 1.* First, by using the rightmost representation in (14), the existence of a minimizer  $g_X$  can be shown as in the proof of Theorem 2 when replacing (A.4) with the following, more general argument. Our assumption that  $\mathbb{E}[\gamma(X)|w(X)]$  is finite yields with Fatou's lemma that

$$\mathbb{E}[\gamma(X)g_0(X)] \geq \limsup_{n \uparrow \infty} \mathbb{E}[\gamma(X)\tilde{g}_n(X)] \geq \liminf_{n \uparrow \infty} \mathbb{E}[\gamma(X)g_n(X)] \geq x_0.$$

Next, we will use the identity

$$\text{ES}_p(Y) = \min_{z \in \mathbb{R}} \left( \frac{1}{1-p} \mathbb{E}[(Y+z)_+] - z \right), \quad Y \in L^1,$$

where the minimum is attained at  $z = F_Y^{-1}(1-p)$ ; see, e.g., Proposition 4.51 in Föllmer and Schied (2016) and note that the proof given there works without modification for  $Y \in L^1$  and does not require the assumption  $Y \in L^\infty$ . Thus, we are in the setting of Theorem 2, and the remainder of the present proof follows exactly as for that result. All one needs to note in addition is that  $\text{ES}_p$  is continuous on  $L^1$ .  $\square$

*Proof of Theorem 3.* Let  $\ell^*(y) := \sup_{x \in \mathbb{R}}(xy - \ell(x))$  be the Fenchel–Legendre transform of  $\ell$ . The risk measure  $\rho$  can be represented in the form

$$\rho(Y) = \max_{\mathbb{Q} \ll \mathbb{P}} \left( \mathbb{E}_{\mathbb{Q}}[Y] - \inf_{\lambda > 0} \frac{1}{\lambda} \left( x_0 + \mathbb{E} \left[ \ell^* \left( \lambda \frac{d\mathbb{Q}}{d\mathbb{P}} \right) \right] \right) \right), \quad Y \in L^\infty; \quad (\text{A.6})$$

see Theorem 4.115 in [Föllmer and Schied \(2016\)](#). Using this representation, the existence of a minimizer  $g_X \in \mathcal{G}$  is established as in the proof of Theorem 2.

As shown at the beginning of the proof of Proposition 4.113 in [Föllmer and Schied \(2016\)](#),  $z^* := \rho(g_X(X))$  is the unique solution of the equation  $\mathbb{E}[\ell(g_X(X) - z)] = x_0$ . It follows from here that  $g_X$  minimizes  $\mathbb{E}[\ell(g(X) - z^*)]$  over  $g \in \mathcal{G}$ . Indeed, suppose by way of contradiction that there is  $g_0 \in \mathcal{G}$  for which  $\mathbb{E}[\ell(g_0(X) - z^*)] < \mathbb{E}[\ell(g_X(X) - z^*)]$ . Then the solution,  $z_0 = \rho(g_0(X))$ , of the equation  $\mathbb{E}[\ell(g_0(X) - z)] = x_0$  will be strictly smaller than  $z^*$ , a contradiction to the optimality of  $g_X$ . The proof of Theorem 4 thus yields the structure of  $g_X$  as a  $\mu_X$ -a.e. continuous function. The robustness of  $\rho$  now follows as in the proof of Theorem 4 by using the fact that  $\rho$  enjoys the so-called Lebesgue property, which in turn is a consequence of Exercise 4.2.3 and Proposition 4.113 in [Föllmer and Schied \(2016\)](#).  $\square$

*Proof of Corollary 2.* Let  $\rho(Y)$  denote the expectile of  $Y \in L^1$  and  $\ell$  the convex loss function  $\ell(x) = \tau x_+ - (1 - \tau)x_-$ . We have  $\ell^*(y) = 0$  if  $1 - \tau \leq y \leq \tau$  and  $\ell^*(y) = +\infty$  otherwise. Hence, letting

$$\tilde{\rho}(Y) := \sup \left\{ \mathbb{E}_{\mathbb{Q}}[Y] : \mathbb{Q} \ll \mathbb{P} \text{ and there exists } \lambda > 0 \text{ s.t. } 1 - \tau \leq \lambda \frac{d\mathbb{Q}}{d\mathbb{P}} \leq \tau \right\}, \quad Y \in L^1, \quad (\text{A.7})$$

the identity (A.6) yields that  $\rho(Y) = \tilde{\rho}(Y)$  for  $Y \in L^\infty$ . For  $Y \in L^1$  and  $n \in \mathbb{N}$ , we let  $Y_n := (-n) \vee Y \wedge n$ . Then we have  $\rho(Y_n) = \tilde{\rho}(Y_n)$ . It is easy to see that  $\rho(Y_n) \rightarrow \rho(Y)$ . Moreover, the set of densities of those probability measure  $\mathbb{Q} \ll \mathbb{P}$  for which there exists  $\lambda > 0$  such that  $1 - \tau \leq \lambda d\mathbb{Q}/d\mathbb{P} \leq \tau$  is bounded in  $L^\infty$ . Therefore, Theorem 4.2 in [Cheridito and Li \(2009\)](#) implies that  $\tilde{\rho}$  is norm continuous on  $L^1$ . Consequently,  $\tilde{\rho}(Y_n) \rightarrow \tilde{\rho}(Y)$ , and we conclude that  $\tilde{\rho}(Y) = \rho(Y)$  for all  $Y \in L^1$ . Using this identity and the norm continuity of  $\rho$ , the robustness of  $\rho$  now follows as in the proof of Theorem 3.  $\square$

*Proof of Proposition 4.* The result follows from Theorem 8.26 of [Föllmer and Schied \(2016\)](#). The fact that  $r$  is a  $p$ -quantile of  $g_X(X)$  is stated in the proof of Theorem 8.26 in [Föllmer and Schied \(2016\)](#).  $\square$

## A.4 Proofs in Section 6

*Proof of Theorem 4.* Consider the function

$$\ell^*(z, x) := \sup_{v(x) \leq y \leq w(x)} (yz - \ell(y)), \quad (\text{A.8})$$

defined for  $y \in \mathbb{R}$  and  $x \in \mathbb{R}^n$ . Let  $y^*(x, z)$  denote the largest maximizer. We must have

$$\begin{aligned} y^*(x, z) = v(x) &\iff \ell'_-(y) > z \quad \text{for all } y \in (v(x), w(x)], \\ y^*(x, z) = w(x) &\iff \ell'_-(y) \leq z \quad \text{for all } y \in (v(x), w(x)]. \end{aligned} \quad (\text{A.9})$$

Moreover,  $\ell'_-(y^*(x, z)) \leq z \leq \ell'_+(y^*(x, z))$  if  $y^*(x, z) \in (v(x), w(x))$  (see, e.g., Proposition A.9 (a) in Föllmer and Schied (2016)). Letting  $I(z) := \inf\{y : \ell'_-(y) > z\} = \sup\{y : \ell'_-(y) \leq z\}$  denote the right-continuous generalized inverse function of  $\ell'_-$ , we hence see that  $y^*(x, z) = I(z)$  in the latter case. Altogether, we obtain that  $y^*(x, z) = v(x) \vee I(z) \wedge w(x)$ .

Let us define

$$g^{(c)}(x) := v(x) \vee I(c\gamma(x)) \wedge w(x), \quad x \in \mathbb{R}^n, \quad c \in \mathbb{R}. \quad (\text{A.10})$$

The function  $I$  is nondecreasing and hence has at most countably many jumps, which form a  $\mu_X \circ \gamma^{-1}$ -nullset, due to our Assumption P. Due to our assumption  $\mathbb{E}[\gamma(X)|w(X)] < \infty$ , we may apply the monotone convergence theorem, which yields that the function  $c \mapsto \mathbb{E}[\gamma(X)g^{(c)}(X)]$  decreases continuously from  $\mathbb{E}[\gamma(X)w(X)] > x_0$  to  $K := \mathbb{E}[\gamma(X)(v(X) \vee I(0) \wedge w(X))]$  as  $c$  decreases from  $+\infty$  to 0. Let us first consider the case in which  $K < x_0$ . In this case, there is some  $c^* > 0$  for which  $\mathbb{E}[\gamma(X)g^{(c^*)}(X)] = x_0$ . We show now that  $g_X := g^{(c^*)}$  is optimal. Indeed, from (A.8) and our definition of  $g_X$ , it is clear that for arbitrary  $g \in \mathcal{G}$ ,

$$c^*\gamma(X)g_X(X) - \ell(g_X(X)) = \ell^*(c^*\gamma(X), X) \geq c^*\gamma(X)g(X) - \ell(g(X)). \quad (\text{A.11})$$

Taking expectations on both sides of (A.11) and using that  $\mathbb{E}[\gamma(X)g(X)] \geq x_0$  hence yields that  $\mathbb{E}[\ell(g_X(X))] \leq \mathbb{E}[\ell(g(X))]$ , which is the desired optimality.

Let us now turn to the case in which  $K \geq x_0$ . To this end, consider  $a := \inf_y \ell'_-(y) \geq 0$  and  $b := \sup_y \ell'_-(y) \in [a, \infty]$ . Then  $I(z) = -\infty$  for  $z < a$  and  $I(z) = +\infty$  for  $z \geq b$ . Moreover,  $I(a) = \lim_{z \downarrow a} I(z)$  is finite if and only if  $\ell$  is linear on  $(-\infty, I(a)]$  with slope  $a$ . Since  $K \geq x_0$  can only occur if  $I(0)$  is finite and we clearly have  $I(0) \leq I(a)$ , it follows that  $\ell$  is linear on  $(-\infty, I(a)]$  and  $I(0) = I(a)$ . On the other hand, the slope of  $\ell$  on  $(I(a), \infty)$  will be greater than  $a$ . Therefore, any function  $g \in \mathcal{G}$  taking values greater than  $v \vee I(a)$  with positive  $\mu_X$ -probability

must be suboptimal, provided that we can solve the following auxiliary problem:

$$\text{minimize } \mathbb{E}[\ell(g(X))] \text{ over } g \in \mathcal{G}_n \text{ with } v \leq g \leq v \vee I(a) \wedge w \text{ and } \mathbb{E}[\gamma(X)g(X)] \geq x_0. \quad (\text{A.12})$$

If  $a = 0$ , so that  $\ell$  is flat on  $(-\infty, I(a)]$ , then every  $g$  satisfying the constraints in (A.12) will be optimal. For instance, we can take

$$g_X := f \quad \text{for } f := v \vee I(a) \wedge w. \quad (\text{A.13})$$

If  $a > 0$ , then we let  $h := f - v$  and replace  $g$  in (A.12) with  $f - g$ . Then (A.12) is equivalent to the auxiliary problem,

$$\text{maximize } \mathbb{E}[g(X)] \text{ over } g \in \mathcal{G}_n \text{ with } 0 \leq g \leq h \text{ and } \mathbb{E}[\gamma(X)g(X)] \leq K - x_0. \quad (\text{A.14})$$

If  $K = x_0$ , this problem has only the trivial solution  $g \equiv 0$ , and so (A.13) is clearly the  $\mu_X$ -a.s. unique solution to (A.12). For  $K > x_0$ , we choose  $c_0 > 0$  such that  $\mathbb{E}[\gamma(X)\mathbb{1}_{\{\gamma(X) \leq c_0\}}h(X)] = K - x_0$ ; this is possible, because, by way of the linearity of  $\ell$  on  $(-\infty, I(a)]$ , our assumption that both  $\mathbb{E}[\ell(v(X))]$  and  $\mathbb{E}[\gamma(X)|w(X)]$  are finite implies that  $\mathbb{E}[\gamma(X)\mathbb{1}_{\{\gamma(X) \leq c\}}h(X)]$  is a finite and continuous function of  $c \in \mathbb{R}$ . Now define  $g^* := h\mathbb{1}_{\{\gamma \leq c_0\}}$  and take any other  $g \in \mathcal{G}_n$  satisfying the constraints in (A.14). Then we have  $(\gamma - c_0)(g - g^*) \geq 0$  and hence

$$0 \leq \mathbb{E}[(\gamma(X) - c_0)(g(X) - g^*(X))] \leq -c_0(\mathbb{E}[g(X)] - \mathbb{E}[g^*(X)]).$$

This shows that  $g^*$  solves (A.14). It follows that

$$g_X := f - g^* = (v \vee I(a) \wedge w)\mathbb{1}_{\{\gamma > c_0\}} + v\mathbb{1}_{\{\gamma \leq c_0\}} \quad (\text{A.15})$$

solves our original problem in case  $K > x_0$ .

To summarize, our original optimization problem admits a solution  $g_X$  that has one of the forms (A.10), (A.13), or (A.15). With this minimizer at hand, we can now proceed to prove the asserted robustness. So suppose that  $Z_k \rightarrow X$  in  $L_n^{pq^+}$ . Since the functions  $v$ ,  $w$ , and  $\gamma$  are continuous  $\mu_X$ -a.e. and since  $I$  has at most countably many discontinuities, we have  $\ell(g_X(Z_k)) \rightarrow \ell(g_X(X))$  in  $L^0$ . Moreover,

$$\begin{aligned} |\ell(g_X(Z_k))| &\leq |\ell_-(v(Z_k))| + |\ell_+(w(Z_k))| \\ &\leq c_1(1 + |Z_k|^{rq^-}) + c_2(1 + |Z_k|^{pq^+}) \leq c_3(1 + |Z_k|^{pq^+}). \end{aligned}$$



It follows that the sequence  $|\ell(g_X(Z_k))|$  is uniformly integrable and so  $\mathbb{E}[\ell(g_X(Z_k))] \rightarrow \mathbb{E}[\ell(g_X(X))]$ . This is the asserted robustness.  $\square$

## A.5 Proofs in Section 9

To prove Proposition 5, we need the following two lemmas. In what follows, we denote by  $A_\varepsilon = \{x \in \mathbb{R} : |x - y| \leq \varepsilon \text{ for some } y \in A\}$  for a set  $A \subset \mathbb{R}$  and  $\varepsilon > 0$ .

**Lemma A.2.** *If  $A \subset \mathbb{R}$  is either compact or an interval, then  $A_\varepsilon \in \mathcal{B}(\mathbb{R})$  and*

$$\sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \mathbb{P}(Y \in A) = \mathbb{P}(X \in A_\varepsilon).$$

*Proof.* If  $A$  is a compact set or an interval, then so is  $A_\varepsilon$ , which proves  $A_\varepsilon \in \mathcal{B}(\mathbb{R})$ . Next, for any  $Y \in L^\infty$  with  $\pi_1^\infty(Y, X) \leq \varepsilon$ , the condition  $Y \in A$  implies  $X \in A_\varepsilon$  a.s. Therefore,  $\mathbb{P}(Y \in A) \leq \mathbb{P}(X \in A_\varepsilon)$ , leading to  $\sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \mathbb{P}(Y \in A) \leq \mathbb{P}(X \in A_\varepsilon)$ . To show the opposite direction of the inequality, it suffices to take  $Y = f_A(X)\mathbf{1}_{\{X \in A_\varepsilon\}} + X\mathbf{1}_{\{X \notin A_\varepsilon\}}$ , where, for a compact set  $A$ ,  $f_A(x)$  is a nearest point of  $x$  in  $A$  (to be precise, there can be two such nearest points; by taking  $f_A(x)$  to be the lower of the two,  $f_A$  becomes lower semicontinuous and, hence, measurable). In the case in which  $A$  is a nondegenerate interval, we fix a point  $a$  in the interior of the interval and let

$$f_A(x) = \begin{cases} a \vee (x - \varepsilon) & \text{if } x \geq \sup A, \\ a \wedge (x + \varepsilon) & \text{if } x \leq \inf A, \\ x & \text{otherwise.} \end{cases}$$

In both cases,  $|Y - X| \leq \varepsilon$ , and  $\mathbb{P}(Y \in A) = \mathbb{P}(X \in A_\varepsilon)$ , leading to the desired result.  $\square$

**Lemma A.3.** *Let  $\varepsilon > 0$ ,  $p \in [1/2, 1)$ , and suppose that  $X$  satisfies Assumption D. If  $A \subset \mathbb{R}$  is a compact set or an interval satisfying  $\sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \mathbb{P}(Y \in A) \leq 1 - p$ , then*

$$\mathbb{P}(X > \text{VaR}_p(X) + \varepsilon) \geq \mathbb{P}(X \in A).$$

*Proof.* By letting  $A^* = (\text{VaR}_p(X) + \varepsilon, \infty)$ , the assertion can be rewritten as  $\mathbb{P}(X \in A^*) \geq \mathbb{P}(X \in A)$ . By Lemma A.2, we have

$$1 - p = \mathbb{P}(X \in A_\varepsilon^*) = \sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \mathbb{P}(Y \in A^*). \quad (\text{A.16})$$

If  $\mathbb{P}(X \in A_\varepsilon) < 1 - p$ , we can enlarge  $A$  to obtain  $\mathbb{P}(X \in A_\varepsilon) = 1 - p$ . Then  $x := \inf(A_\varepsilon)$  satisfies  $x \leq \text{VaR}_p(X)$  since  $\mathbb{P}(X \leq x) \leq 1 - \mathbb{P}(X \in A_\varepsilon) = p$ .

We consider two cases separately. First, we assume  $x > \text{ess-inf}X$ . It is from the definition of  $x$  that  $\inf(A) = x + \varepsilon$ . Hence,  $(x, x + \varepsilon) \subset A_\varepsilon \setminus A$ . Also note that  $\mathbb{P}(X \in (x, x + \varepsilon)) \geq \mathbb{P}(X \in (\text{VaR}_p(X), \text{VaR}_p(X) + \varepsilon))$  since  $X$  has a decreasing density and  $x \leq \text{VaR}_p(X)$ . Therefore, we have

$$\begin{aligned} \mathbb{P}(X \in A) &= \mathbb{P}(X \in A_\varepsilon) - \mathbb{P}(X \in A_\varepsilon \setminus A) \\ &\leq 1 - p - \mathbb{P}(X \in (x, x + \varepsilon)) \\ &\leq 1 - p - \mathbb{P}(X \in (\text{VaR}_p(X), \text{VaR}_p(X) + \varepsilon)) = \mathbb{P}(X \in A^*). \end{aligned}$$

Next, we assume  $x \leq \text{ess-inf}X$ . Since  $p \in [1/2, 1)$ , we have  $\mathbb{P}(X < \text{VaR}_p(X) + \varepsilon) > p \geq 1 - p$ . Because  $\mathbb{P}(X \in A_\varepsilon) = 1 - p$  and  $x \leq \text{ess-inf}X$ , there exists  $x_0 \in (x, \text{VaR}_p(X) + \varepsilon)$  such that  $x_0 \notin A_\varepsilon$ . Let  $x_1 = \sup\{y < x_0 : y \in A_\varepsilon\}$ . By the definition of  $A_\varepsilon$  and  $x_1$ , we have  $x_1 - \varepsilon > x$  and  $(x_1 - \varepsilon, x_1) \subset A_\varepsilon \setminus A$ . Using a similar argument as in the first case, we have

$$\begin{aligned} \mathbb{P}(X \in A) &= \mathbb{P}(X \in A_\varepsilon) - \mathbb{P}(X \in A_\varepsilon \setminus A) \\ &\leq 1 - p - \mathbb{P}(X \in (x_1 - \varepsilon, x_1)) \\ &\leq 1 - p - \mathbb{P}(X \in (\text{VaR}_p(X), \text{VaR}_p(X) + \varepsilon)) = \mathbb{P}(X \in A^*). \end{aligned}$$

We conclude that, in both cases,  $\mathbb{P}(X \in A^*) \geq \mathbb{P}(X \in A)$ . □

*Proof of Proposition 5.* Recall that  $\mathcal{G}$  is given by (20), and  $\mathbb{E}_X[h]$  means  $\mathbb{E}[h(X)]$  for any function  $h$ . Take an arbitrary  $g \in \mathcal{G}$ . Denote by

$$a = \sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(g(Y)),$$

and let  $h$  be given by

$$h(x) = m\mathbf{1}_{\{g(x) > a\}} + a\mathbf{1}_{\{g(x) \leq a\}}, \quad x \in \mathbb{R}. \quad (\text{A.17})$$

For all  $Y \in L^\infty$  with  $\pi_1^\infty(Y, X) \leq \varepsilon$ , we have  $\text{VaR}_p(g(Y)) \leq a$ . Therefore,  $\mathbb{P}(g(Y) > a) \leq 1 - p$ , which implies  $\text{VaR}_p(h(Y)) \leq a$ . Thus,

$$\sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(h(Y)) = a.$$

Note that  $h \in \mathcal{G}$  since  $m \geq h(x) \geq g(x) \geq 0$ ,  $x \in \mathbb{R}$ . Therefore, for any  $g \in \mathcal{G}$ , we can find some

$h \in \mathcal{G}$  of the form (A.17), such that

$$\sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(h(Y)) = \sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(g(Y)).$$

As a consequence, it suffices to search for optimizers  $h \in \mathcal{G}$  of the form (A.17). Moreover, for such an  $h$ , we have  $\mathbb{E}_X[\gamma h] = m\mathbb{Q}(g(X) > a) + a\mathbb{Q}(g(X) \leq a)$ , where  $\mathbb{Q}$  is given by  $d\mathbb{Q}/d\mathbb{P} = \gamma$ . Due to the inner regularity of the law  $\mathbb{Q} \circ X^{-1}$ , we can find, for any  $a' > a$ , a compact set  $K \subset \{g(X) > a\}$  such that  $h'(x) = m\mathbb{1}_{\{x \in K\}} + a'\mathbb{1}_{\{x \in K^c\}}$  satisfies  $\mathbb{E}_X[\gamma h'] \geq \mathbb{E}_X[\gamma h] \geq x_0$ . Since  $\mathbb{P}(Y \in K) \leq \mathbb{P}(g(Y) > a) \leq 1 - p$  for all  $Y \in L^\infty$  with  $\pi_1^\infty(Y, X) \leq \varepsilon$ , we have

$$\sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(h'(Y)) = a'.$$

Let us denote by  $\mathcal{K}$  the class of all compact set  $K \subset \mathbb{R}$  satisfying  $\mathbb{P}(Y \in K) \leq 1 - p$  for all  $Y \in L^\infty$  with  $\pi_1^\infty(Y, X) \leq \varepsilon$ . The above argument shows that  $\mathcal{K}$  is not empty. Define a function

$$h_K(x) = m\mathbb{1}_{\{x \in K\}} + a_K\mathbb{1}_{\{x \in K^c\}}, \quad x \in \mathbb{R}, \quad (\text{A.18})$$

where  $a_K \in \mathbb{R}$  is such that  $\mathbb{E}_X[\gamma h_K] = x_0$ . The existence of  $a_K$  is guaranteed by  $\mathbb{P}(X \in K^c) \geq p > 0$ . Note that  $0 < a_K < m$  since  $x_0 < m$  and  $q > 0$ .

The preceding argument shows that it is sufficient to construct a function  $h^*$  such that  $\mathbb{E}[\gamma h^*(X)] = x_0$  and

$$\sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(h^*(Y)) \leq \sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(h_K(Y)) = a_K, \quad (\text{A.19})$$

for all  $K \in \mathcal{K}$ . We define  $h^*$  by

$$h^*(x) = m\mathbb{1}_{\{x > c + \varepsilon\}} + a^*\mathbb{1}_{\{x \leq c + \varepsilon\}}, \quad x \in \mathbb{R},$$

where  $a^* \leq m$  is such that  $\mathbb{E}_X[\gamma h^*] = x_0$ . Now let  $K \in \mathcal{K}$  be given and  $h_K$  of the form (A.18). We take  $k \in \mathbb{R}$  such that  $\mathbb{P}(X > k) = \mathbb{P}(X \in K)$ . Lemma A.3 gives

$$\mathbb{P}(X > c + \varepsilon) \geq \mathbb{P}(X \in K) = \mathbb{P}(X > k)$$

and hence  $k \geq c + \varepsilon$ . Moreover, since  $\gamma$  is an increasing function of  $X$ , the upper Hardy–Littlewood inequality, in the form of Föllmer and Schied (2016, Theorem A.28), yields

$$x_0 \leq \mathbb{E}_X[\gamma h_K] \leq \mathbb{E}[\gamma(X)(m\mathbb{1}_{\{X > k\}} + a_K\mathbb{1}_{\{X \leq k\}})] \leq \mathbb{E}[\gamma(X)(m\mathbb{1}_{\{X > c + \varepsilon\}} + a_K\mathbb{1}_{\{X \leq c + \varepsilon\}})].$$

Our condition  $\mathbb{E}_X[\gamma h^*] = x_0$  therefore yields  $a_K \geq a^*$ . Since, moreover, by construction,

$$\sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(h^*(Y)) = a^* \quad \text{and} \quad \sup_{\pi_1^\infty(Y, X) \leq \varepsilon} \text{VaR}_p(h_K(Y)) = a_K,$$

we conclude that (A.19) holds and that  $h^*$  is hence a solution to Problem (21).  $\square$

## A.6 Robustness of VaR and ES in an unbounded setting

For  $\rho = \text{VaR}_p$  or  $\rho = \text{ES}_p$ , we consider the unbounded optimization problem

$$\text{to minimize: } \rho(g(X)) \quad \text{subject to } g \in \mathcal{G}_n, \quad \mathbb{E}[\gamma(X)g(X)] \geq x_0, \quad (\text{A.20})$$

where  $\gamma : \mathbb{R}^n \rightarrow (0, \infty)$  and  $x_0 \in \mathbb{R}$ . Problem (A.20) corresponds to (8) with  $w = \infty$  and  $v = -\infty$ .

**Proposition A.1.** *Assume that  $X$  is continuously distributed,  $\mathbb{E}[\gamma(X)] < \infty$ , and  $p \in (0, 1)$ .*

- (i) *For  $\rho = \text{VaR}_p$ , Problem (A.20) has no solution.*
- (ii) *For  $\rho = \text{ES}_p$ , Problem (A.20) admits a solution if and only if*

$$\text{ess-sup}\gamma(X) \leq \frac{1}{1-p}. \quad (\text{A.21})$$

Moreover, if (A.21) holds, a solution to (A.20) is given by the constant function  $g_X(\cdot) = x_0$ .

*Proof.* Denote by

$$\mathcal{G}_{\text{ub}} = \{g \in \mathcal{G}_n : \mathbb{E}[\gamma(X)g(X)] \geq x_0\}.$$

- (i) Let  $A$  be a set such that  $\mathbb{P}(X \in A) = 1 - p$ . Write  $\lambda = \mathbb{E}[\gamma(X)\mathbb{1}_{\{X \in A\}}] > 0$ . For  $d < x_0$ , define the function

$$g_d(x) = d + \frac{x_0 - d}{\lambda} \mathbb{1}_{\{x \in A\}}, \quad x \in \mathbb{R}^n.$$

Clearly,  $g_d(X) \in \mathcal{G}_{\text{ub}}$  because  $\mathbb{E}[\gamma(X)g_d(X)] = d + \frac{x_0 - d}{\lambda} \mathbb{E}[\gamma(X)\mathbb{1}_{\{X \in A\}}] = x_0$ . On the other hand,  $\text{VaR}_p(g_d(X)) = d$ . Letting  $d \rightarrow -\infty$ ,

$$\text{VaR}_p(X; \mathcal{G}_{\text{ub}}) = \inf\{\rho(\text{VaR}_p(g(X))) : g \in \mathcal{G}_{\text{ub}}\} = -\infty,$$

and hence (A.20) does not have an optimizer.

- (ii) By the dual representation of  $\text{ES}_p$  in (14), we have

$$\text{ES}_p(Y) = \sup_{B \in \mathcal{B}_p} \mathbb{E}[BY] \quad \text{for all } Y \in L^1,$$

where  $\mathcal{B}_p = \{B \in L^\infty : \mathbb{E}[B] = 1, 0 \leq B \leq \frac{1}{1-p}\}$ . If  $\text{ess-sup}\gamma(X) \leq \frac{1}{1-p}$ , then  $\gamma(X) \in \mathcal{B}_p$ , and hence for any  $g \in \mathcal{G}_{\text{ub}}$ ,  $\text{ES}_p(g(X)) \geq \mathbb{E}[\gamma(X)g(X)] \geq x_0$ . Clearly, taking the constant function  $g_X(\cdot) = x_0$  we have  $g_X \in \mathcal{G}_{\text{ub}}$  and  $\text{ES}_p(g_X(X)) = x_0$ . Therefore,  $g_X$  is a solution to Problem (A.20).

Next, assume  $\text{ess-sup}\gamma(X) > \frac{1}{1-p}$ . Denote by  $y = \mathbb{E}[\gamma(X)\mathbb{1}_{\{\gamma(X) > \frac{1}{1-p}\}}] > 0$  and  $k = \text{ES}_p(\mathbb{1}_{\{\gamma(X) > \frac{1}{1-p}\}})$ . Note that  $k \leq y$  because

$$\begin{aligned} \text{ES}_p\left(\mathbb{1}_{\{\gamma(X) > \frac{1}{1-p}\}}\right) &= \sup_{B \in \mathcal{B}_p} \mathbb{E}[B\mathbb{1}_{\{\gamma(X) > \frac{1}{1-p}\}}] \\ &\leq \frac{1}{1-p} \mathbb{E}[\mathbb{1}_{\{\gamma(X) > \frac{1}{1-p}\}}] < \mathbb{E}[\gamma(X)\mathbb{1}_{\{\gamma(X) > \frac{1}{1-p}\}}]. \end{aligned}$$

For  $\lambda > 0$ , take  $g_\lambda(x) = \lambda\mathbb{1}_{\{\gamma(x) > \frac{1}{1-p}\}} - \lambda y + x_0$ ,  $x \in \mathbb{R}^n$ . It is clear that  $\mathbb{E}[\gamma(X)g_\lambda(X)] = \lambda y - \lambda y + x_0 = x_0$ , and hence  $g_\lambda \in \mathcal{G}_{\text{ub}}$ . We can calculate

$$\text{ES}_p(g_\lambda(X)) - \mathbb{E}[\gamma(X)g_\lambda(X)] = \lambda \left( \text{ES}_p\left(\mathbb{1}_{\{\gamma(X) > \frac{1}{1-p}\}}\right) - y \right) = \lambda(k - y).$$

Letting  $\lambda \rightarrow \infty$ , we get

$$\inf\{\text{ES}_p(g(X)) : g \in \mathcal{G}_{\text{ub}}\} = -\infty,$$

and hence there is no solution to Problem (A.20).  $\square$

As a direct consequence of Proposition A.1, for any choice of  $(\mathcal{Z}, \pi)$ ,  $\text{VaR}_p$  is not robust against optimization for  $(\mathcal{G}_{\text{ub}}, \mathcal{Z}, \pi)$ , and if (A.21) holds, then  $\text{ES}_p$  is robust against optimization for  $(\mathcal{G}_{\text{ub}}, \mathcal{Z}, \pi)$ . Clear from the proof, the assumption that  $X$  has a continuous distribution is only used in part (i), and it can be relaxed to requiring  $\mathbb{P}(X \in A) \in (0, 1-p)$  for some event  $A$ .