



## Decision Support

## Robust and Pareto optimality of insurance contracts

Alexandru V. Asimit<sup>a,\*</sup>, Valeria Bignozzi<sup>b</sup>, Ka Chun Cheung<sup>c</sup>, Junlei Hu<sup>a</sup>, Eun-Seok Kim<sup>d</sup><sup>a</sup> Cass Business School, City University London, London EC1Y 8TZ, United Kingdom<sup>b</sup> Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy<sup>c</sup> Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong<sup>d</sup> Department of Management, Leadership and Organisations, Middlesex University, London NW4 4BT, United Kingdom

## ARTICLE INFO

## Article history:

Received 2 September 2016

Accepted 12 April 2017

Available online 20 April 2017

## Keywords:

Uncertainty modelling

Linear programming

Robust/Pareto optimal insurance

Risk measure

Robust optimisation

## ABSTRACT

The optimal insurance problem represents a fast growing topic that explains the most efficient contract that an insurance player may get. The classical problem investigates the ideal contract under the assumption that the underlying risk distribution is known, i.e. by ignoring the parameter and model risks. Taking these sources of risk into account, the decision-maker aims to identify a robust optimal contract that is not sensitive to the chosen risk distribution. We focus on *Value-at-Risk* (VaR) and *Conditional Value-at-Risk* (CVaR)-based decisions, but further extensions for other risk measures are easily possible. The *Worst-case scenario* and *Worst-case regret* robust models are discussed in this paper, which have been already used in robust optimisation literature related to the investment portfolio problem. Closed-form solutions are obtained for the VaR Worst-case scenario case, while *Linear Programming* (LP) formulations are provided for all other cases. A caveat of robust optimisation is that the optimal solution may not be unique, and therefore, it may not be economically acceptable, i.e. Pareto optimal. This issue is numerically addressed and simple numerical methods are found for constructing insurance contracts that are Pareto and robust optimal. Our numerical illustrations show weak evidence in favour of our robust solutions for VaR-decisions, while our robust methods are clearly preferred for CVaR-based decisions.

© 2017 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

Finding the optimal insurance contract has represented a topic of interest in the actuarial science and insurance literature for more than 50 years. The seminal papers of Borch (1960) and Arrow (1963) had opened this field of research and since then, many papers discussed this problem under various assumptions on the risk preferences of the insurance players involved in the contract and how the cost of insurance (known as *premium*) is quantified. Specifically, the optimal contracts in the context of Expected Utility Theory are investigated amongst others in Kaluszka (2005), Kaluszka and Okolewski (2008) and Cai and Wei (2012). Extensive research has been made when the preferences are made via coherent risk measures (as defined in Artzner, Delbaen, Eber, and Heath, 1999; recall that CVaR is an element of this class) and VaR; for example, see Cai and Tan (2007), Balbás, Balbás, and Heras (2009);

2011), Asimit, Badescu, and Verdonck (2013b), Cheung, Sung, Yam, and Yung (2014) and Cai and Weng (2016) among others.

The choice of a risk measure is usually subjective, but VaR and CVaR represent the most known risk measures used in the insurance industry. Solvency II and Swiss Solvency Test are the regulatory regimes for all (re)insurance companies that operate within the European Union and Switzerland, respectively, and their capital requirements are solely based on VaR and CVaR. For these reasons and not only, these standard risk measures have received special attention by academics, practitioners and regulators, and therefore, vivid discussions have risen among them. VaR is criticised for its lack of sub-additivity and it may create regulatory arbitrage in an insurance group (see Asimit, Badescu, & Tsanakas, 2013a). A detailed discussion on possible regulatory arbitrages in a CVaR-based regime is provided in Koch-Medina and Munari (2016). A desirable property for a risk measure is the *elicibility*, which allows one to compare competitive forecasting methods, a property that VaR does have (see Gneiting, 2011). The lack of elicibility for CVaR has been adjusted via the *joint elicibility*, concept formalised in Fissler and Ziegel (2016), but earlier flagged out by Acerbi and Székely (2014). Robustness properties of a risk measure are also of great interest since they imply that the estimate is insensitive to

\* Corresponding author.

E-mail addresses: [asimit@city.ac.uk](mailto:asimit@city.ac.uk) (A.V. Asimit), [valeria.bignozzi@unimib.it](mailto:valeria.bignozzi@unimib.it) (V. Bignozzi), [kccg@hku.hk](mailto:kccg@hku.hk) (K.C. Cheung), [Junlei.Hu.1@city.ac.uk](mailto:Junlei.Hu.1@city.ac.uk) (J. Hu), [E.Kim@mdx.ac.uk](mailto:E.Kim@mdx.ac.uk) (E.-S. Kim).

data contamination. *Parameter risk* (uncertainty with parameter estimation) and *model risk* (uncertainty with model selection) are the two main sources of uncertainty in modelling. The robust statistic has its roots in the papers of Huber (1964) and Hampel (1968), which has been shown to be less appropriate in the context of risk management (see for example, Cont, Deguest, & Scandolo, 2010). A more informative discussion is given in the next section due to its length. Finally, a summary of all properties exhibited by the two risk measures is detailed in the comprehensive work of Emmer, Kratz, and Tasche (2015), but the general conclusion is that there is no evidence for global advantage of one risk measure against the other.

Whenever the model and parameter risks are present, it is prudent to consider insurance contracts that are optimal under a set of plausible models and this is precisely what *robust optimisation* does. It is a vast area of research with applications in various fields and a standard reference is Ben-Tal, El Ghaoui, and Nemirovski (2009), while comprehensive surveys can be found in Ben-Tal and Nemirovski (2008), Bertsimas, Brown, and Caramanis (2011) and Gabrel, Murat, and Thiele (2014).

The aim of the paper is to identify the optimal insurance contract under the model/parameter risk in the robust optimisation sense and understand how robust these solutions are from the practical point of view. That is, we aim to explain how large the uncertainty set should be for relatively small or medium sized historical data sets as is expected in insurance practice. At the same time, since the insurance contract is in fact a risk allocation, it is of great interest to find whether or not our robust contracts are Pareto optimal. Robust optimisation may lead to inefficient risk allocations, i.e. not Pareto optimal, which are clearly not acceptable, and special attention is given to this issue by providing a simple methodology to overcome such caveats of robust optimisation. Our numerical illustrations have shown weak evidence in favour of our robust solutions for VaR-based decisions, which is not surprising due to the erratic behaviour of VaR. On the contrary, CVaR-based decisions are more robust via robust optimisation than using statistical methods, which can be explained by the fact that CVaR takes into account some part of the tail risk as opposed to VaR. Either Worst-case scenario or regret robust optimisations is preferred (comparing to the classical statistical methods) for less (statistically) robust risk measures that are purely tail risk measures, where the estimation is based on a small portion of the sample that explains only the tail risk. We also find that the Worst-case optimisation is once again advantageous even for risk measures that are sensitive to the entire sample, i.e. are not only based on the tail risk.

The structure of the paper is as follows: the next section contains the necessary background and the mathematical formulation of our problems, while Sections 3 and 4 investigate the VaR and CVaR-based optimal insurance contracts, but also discuss simple extensions for distortion risk measures when the moral hazard is removed; these robust solutions are further investigated in Section 5 to becoming Pareto optimal as well; extensive numerical examples are elaborated in Section 6, which help in justifying our conclusions summarised in Section 7.

## 2. Background and problem definition

### 2.1. Optimal insurance

An insurance contract represents a risk transfer between two parties, *insurance buyer* (or simply *buyer*) and *insurance seller* (or simply *seller*). When the buyer is also an insurance company, then the transfer becomes a *reinsurance contract* and the seller is called *reinsurer*. Let  $X \geq 0$  be the total amount that the buyer is liable to pay in the absence of any risk transfer. In addition, the

seller agrees to pay  $R[X]$ , the amount by which the entire loss exceeds the buyer's amount,  $I[X]$ , and clearly we have  $I[X] + R[X] = X$ . The most common risk transfers are the Proportional and Stop-loss contracts for which  $I[X] = cX$  (with  $0 \leq c \leq 1$ ) and  $I[X] = \min\{X, M\}$ , respectively. Note that in order to avoid moral hazard issues (both players are incentivised to reduce the overall risk, i.e.  $I$  and  $R$  are non-decreasing functions),  $I, R \in C^{co}$ , where

$$C^{co} = \{f \text{ is non-decreasing} \mid 0 \leq f(x) \leq x, |f(x) - f(y)| \leq |x - y| \text{ for all } x, y \in \mathfrak{R}\}.$$

The comonotonic risk transfers (as defined above) are omnipresent in practice, but it is not always the case and the mathematical formulation of the feasibility set becomes

$$C = \{f \mid 0 \leq f(x) \leq x \text{ for all } x \in \mathfrak{R}\}.$$

Let  $P$  be the insurance premium, and it is further assumed that any feasible contract satisfies  $0 \leq P \leq \bar{P}$ , where  $\bar{P}$  represents a maximal amount of premium that the buyer would accept to pay. If the loss distribution is known, then the premium calculations are possible via certain rules, known as *premium principles*. A concise review of premium principles can be found in Young (2004). Specifically, if  $\mathcal{P}$  is the probability measure for  $X$ , then  $P \geq \omega_0 + (1 + \theta)\mathbb{H}_{\mathcal{P}}(R[X])$ , where  $\omega_0 \geq 0$  represents some fixed/administrative costs,  $\theta \geq 0$  is the *risk loading* parameter/factor, and  $\mathbb{H}$  is a monotone functional on the space of non-negative random variables that depends on the seller's choice of premium principle. The monotonicity property is of practical importance and it means that if two random losses satisfy  $Y \leq Z$ , then  $\mathbb{H}_{\mathcal{P}}(Y) \leq \mathbb{H}_{\mathcal{P}}(Z)$ . A commonly encountered premium principle is the *distortion premium principle* (see Wang, Young, & Panjer, 1997),

$$\mathbb{H}_{\mathcal{P}}(Y) = \int_0^{\infty} g(\mathcal{P}(Y > y)) dy \tag{2.1}$$

for any non-negative loss random variable  $Y$ , where  $g: [0, 1] \rightarrow [0, 1]$  is non-decreasing with  $g(0) = 0$  and  $g(1) = 1$  known as *distortion function*. When the distortion function is taken to be the identity function, we obtain the *expected value premium principle*, which is standard in the insurance industry. The mathematical formulation of the optimal insurance problem becomes

$$\min_{(R,P) \in C^{co} \times \mathfrak{R}} \{\rho_{\mathcal{P}}(X - R[X] + P)\}, \text{ s.t. } \omega_0 + (1 + \theta)\mathbb{H}_{\mathcal{P}}(R[X]) \leq P \leq \bar{P}, \tag{2.2}$$

where  $\rho_{\mathcal{P}}$  is a risk measure chosen by the buyer to order its preferences to risk. As explained in Section 1, it is first assumed in this paper that  $\rho_{\mathcal{P}} \in \{\text{VaR}, \text{CVaR}\}$ . Recall that the lower script  $\mathcal{P}$  indicates the probability measure under which the risk measurement is made. The VaR of a loss variable  $Y$  at a confidence level  $\alpha \in (0, 1)$ , is given by  $\text{VaR}_{\alpha}(Y; \mathcal{P}) = \inf_{y \in \mathfrak{R}} \{\mathcal{P}(Y \leq y) \geq \alpha\}$ . Note that  $\text{VaR}_{\alpha}$  is representable as in (2.1) with  $g(t) = \mathbb{I}_{\{t > 1 - \alpha\}}$ , where  $\mathbb{I}_A$  represents the indicator operator that assigns the value 1 if  $A$  is true and 0 otherwise. The CVaR risk measure is defined in Rockafeller and Uryasev (2000) as follows:

$$\text{CVaR}_{\alpha}(Y; \mathcal{P}) = \inf_{t \in \mathfrak{R}} \left\{ t + \frac{1}{1 - \alpha} \mathbf{E}_{\mathcal{P}}(Y - t)_+ \right\}, \text{ where } (t)_+ = \max(t, 0). \tag{2.3}$$

Alternative representations are known in the literature (see for example, Acerbi & Tasche, 2002) and one of them is as in (2.1) with  $g(t) = \frac{t}{1 - \alpha} \wedge 1$ .

Due to the monotonicity property of VaR, CVaR and the functional  $\mathbb{H}$ , (2.2) becomes much simpler when removing the economic constraint  $P \leq \bar{P}$  and it has been investigated under various sets of assumptions. Recently, Cheung and Lo (2017) included the latter constraint and analytically solved (2.2) for a large class

of premium principles and risk measures, including the class from (2.1).

The existing literature assumes that the loss distribution is certainly known, and as a result, the parameter and model risks are removed. Small and medium sized samples (present in non-high frequency data, as is usually the case in insurance data) raises many questions when estimating any parameter even if the model risk is completely removed, i.e. the chosen model is correct. Large samples are more concerned with the model risk, which can be reduced if the model is carefully selected. Thus, if we know what we need to estimate, for example the optimal objective function value from (2.2), for which its closed-form solution is required, the elicibility (see Gneiting, 2011) of this functional (induced by the optimal objective function value) is the next step in order to compare various models and reduce the model risk. While VaR is elicitable, VaR and CVaR are jointly elicitable, our functional may not be elicitable or impossible to assess the presence of this property, since one has to find a scoring function to measure the estimation error under the plausible models. Therefore, the model selection for VaR/CVaR do not apply for our problem, even though these two simple risk measures are well-accepted as “good” risk measures. Now, even if we can select the “best” possible model that reduces the uncertainty with the optimal objective function value, we in fact solve a secondary problem, since the main purpose of this exercise is to obtain a robust decision (with respect to the insurance contract, i.e.  $R$ ).

Therefore, it would be interesting to identify a more robust optimal insurance contract that would take into account the parameter and/or model error. Thus, we assume that the reference probability measure  $\mathcal{P}$  is unknown and could be one of the  $m$  possible probability measures  $\{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_m\}$ . Consequently, the premium feasibility constraint becomes

$$\omega_0 + (1 + \theta) \cdot \mathbb{H}_{\mathcal{P}_k}(R[X]) \leq P \leq \bar{P}, \quad \text{for any } k \in \mathcal{M} = \{1, 2, \dots, m\}.$$

A prudent and hopefully robust decision is obtained when investigating the worst-case scenario optimisation problem

$$\begin{cases} \min_{(R,P) \in \mathcal{C} \times \mathfrak{N}} \max_{k \in \mathcal{M}} \{\rho_{\mathcal{P}_k}(X - R[X] + P)\}, \\ \text{s.t. } \omega_0 + (1 + \theta) \mathbb{H}_{\mathcal{P}_k}(R[X]) \leq P \leq \bar{P}, \quad \text{for any } k \in \mathcal{M}. \end{cases} \quad (2.4)$$

An alternative prudent decision can be achieved via the worst-case regret optimisation problem

$$\begin{cases} \min_{(R,P) \in \mathcal{C} \times \mathfrak{N}} \max_{k \in \mathcal{M}} \{\rho_{\mathcal{P}_k}(X - R[X] + P) - \rho_k^*\}, \\ \text{s.t. } \omega_0 + (1 + \theta) \mathbb{H}_{\mathcal{P}_k}(R[X]) \leq P \leq \bar{P}, \quad \text{for any } k \in \mathcal{M}, \end{cases} \quad (2.5)$$

where the buyer’s “regret” is measured with respect to some  $m$  benchmark values  $\rho_k^*$ . Naturally, these values are the optimal objective values for the individual models and are variants of (2.2). Specifically,

$$\begin{cases} \rho_k^* = \min_{(R,P) \in \mathcal{C} \times \mathfrak{N}} \{\rho_{\mathcal{P}_k}(X - R[X] + P)\}, \\ \text{s.t. } \omega_0 + (1 + \theta) \mathbb{H}_{\mathcal{P}_k}(R[X]) \leq P \leq \bar{P}, \quad \text{for any } k \in \mathcal{M}. \end{cases} \quad (2.6)$$

These robust representations have been seen before in various ways. The worst-case type decisions were axiomatically investigated by Gilboa and Schmeidler (1989) in the expected utility context. Not surprisingly, the robust optimisation within the Portfolio Theory has its counterpart; among others, see El Ghaoui, Oks, and Oustry (2003), Zhu and Fukushima (2009), Polak, Rogers, and Sweeney (2010), Zymler, Kuhn, and Rustem (2013) and Kakouris and Rustem (2014). The worst-case and worst-case regret CVaR-based decisions in portfolio optimisation are discussed in Huang, Zhu, Fabozzi, and Fukushima (2010). According to our knowledge, the optimal insurance contract problem under parameter/model uncertainty has been investigated only by Balbás, Balbás, Balbás, and Heras (2015), where only the worst-case is investigated

for a large class of risk measures that includes CVaR, but not VaR, and a particular choice of the uncertainty set of probability measures.

We now discuss the choice of the feasibility set, i.e.  $\mathcal{C}$  or  $\mathcal{C}^{co}$ . Note that whenever the risk transfer is made between two large insurance companies, the moral hazard may not be an issue, due to the presence of rating agencies; rating downgrading has a huge negative commercial impact for such insurance companies and thus, moral hazard is less likely to occur. One may also argue that a risk transfer within an insurance group does not necessarily have to exclude the moral hazard due to the common ownership of the buyer and seller. Nevertheless, the insurance regulator requires the insurer buyer to justify the commercial purpose of such a risk transfer. In the absence of distributional uncertainty, there is a huge literature that discusses whether or not the indemnity of an insurance contract should be comonotone, but in general, the conclusion depends on the nature of the underwritten risk. On the other hand, the classical Pareto optimality problem explains the shape of an “optimal” contract and the extensive existing literature discusses how viable the comonotonic property is; an interesting discussion appears in Huberman, Mayers, and Smith (1983). Optimal transfers are shown to be comonotone (for a large class of risk preferences) in Landsberger and Meilijson (1994) if the total risk is finite, while Ludkovski and Rüschenendorf (2008) extends this results to unbounded risks. In summary, choosing between a set of feasible contracts given by  $\mathcal{C}$  or  $\mathcal{C}^{co}$  is related to the specific nature of the total risk that is shared and the insurance players’ risk preferences whenever the total risk distribution is known. In the presence of distributional uncertainty, the choice of feasibility set is sensitive to the nature of the total risk. Therefore, solutions to Problems (2.4)–(2.6) are given to non-comonotone contracts set,  $\mathcal{C}$ , whenever possible, otherwise the comonotone contracts set  $\mathcal{C}^{co}$  is chosen. Recall that we do not intend to characterise the optimal contract, but instead we examine when our proposed robust methods reduce the effect of distributional uncertainty.

Note that the feasible sets of Problems (2.4)–(2.6) are empty if  $\omega_0 > \bar{P}$ . We now gather the set of assumptions, stated as Assumption 2.1, under which the results of the paper hold.

**Assumption 2.1.** We consider  $m$  possible probability models  $\{\mathcal{P}_1, \dots, \mathcal{P}_m\}$  and the reference probability model  $\mathcal{P}$  may or may not belong to this set. Denote  $\mathcal{M} = \{1, \dots, m\}$ . Let  $X \geq 0$  be a loss random variable and denote  $F_k(\cdot) = \mathcal{P}_k(X \leq \cdot)$ ,  $k \in \mathcal{M}$ , its cumulative distribution function (cdf) under  $\mathcal{P}_k$ , we write  $X \sim \mathcal{P}_k$ , and  $\bar{F}_k(\cdot) = 1 - F_k(\cdot)$  its corresponding survival function. Moreover,  $\omega_0 \leq \bar{P}$ . The premium principle is based on a monotonic functional  $\mathbb{H}$ .

## 2.2. Robustness of risk measures

In the last few years there has been a wide and open debate on the robustness properties of VaR and CVaR, with relevant contributions from regulators, practitioners and academics. These risk measures, that we denote for brevity  $\rho$  depend on the probability model  $\mathcal{P}$  used. The key question is whether a small perturbation of the probability model  $\mathcal{P}$  would result in a small perturbation of  $\rho_{\mathcal{P}}$ , which is detailed in the next definition.

**Definition 2.1.** Let  $X_n, n \geq 1$  be a sequence of random variables with distribution  $\mathcal{P}_n$ ,  $n \geq 1$  and  $X$  a random variable with distribution  $\mathcal{P}$ . A risk measure  $\rho_{\mathcal{P}}(X)$  is (statistically) robust at  $\mathcal{P}$ , if  $\lim_{n \rightarrow \infty} d(\mathcal{P}_n, \mathcal{P}) = 0$  implies  $\lim_{n \rightarrow \infty} |\rho_{\mathcal{P}_n}(X_n) - \rho_{\mathcal{P}}(X)| = 0$  for some distance  $d$  between probability measures.

Different specifications of the metric  $d$  correspond to different notions of robustness. For instance, Kiesel, Rühlicke, Stahl, and Zheng (2016) consider the Wasserstein distance

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

under which both VaR and CVaR are robust. Cont et al. (2010) use the Lévy distance and show that there is a partial conflict between coherent risk measures (including CVaR) and Hampel's classical notion of robustness, Krättschmer, Schied, and Zähle (2014) and Delbaen, Bellini, Bignozzi, and Ziegel (2016) consider continuity with respect to the  $\varphi$ -weak topology. We refer the interested reader to Emmer et al. (2015) for a brief summary on the topic. Statistical robustness is particularly relevant when the probability measure is estimated from available data; indeed if the estimated probability measure  $\mathcal{P}_n$  is sufficiently close to the real one (that is  $d(\mathcal{P}_n, \mathcal{P}) \rightarrow 0$ ) and the risk measure is robust, then  $\rho_{\mathcal{P}_n}$  can be considered as a good approximation of  $\rho_{\mathcal{P}}$ .

Due to data scarcity, as it is often the case in practice, the estimates based on the empirical measure exhibit weak statistical evidence and alternative methods are necessary to consider. For example, a more conservative approach is to consider a robustified risk measure  $\bar{\rho}$  defined as follows:

$$\bar{\rho}(X) = \sup_{\mathcal{P} \in \mathcal{S}} \rho_{\mathcal{P}}(X), \tag{2.7}$$

where  $\rho_{\mathcal{P}}(X)$  represents the risk measure for the random loss  $X$  with probability distribution  $\mathcal{P}$  and  $\mathcal{S}$  is a set of candidates models. This approach is not new in the literature, it is at the basis of decision making under ambiguity (that is when there is uncertainty about the probability distribution). The simple idea of this approach is that when there is ambiguity between different models, a conservative and therefore robust approach is to select the one that represents the worst scenario. In Assumption 2.1, we assume that the real probability model  $\mathcal{P}$  may not belong to the set  $\mathcal{S}$ . Indeed, since  $\mathcal{P}$  is unknown, we cannot guarantee that it belongs to the set of models considered. Note that taking the supremum over a set of models reduces the impact of model risk, but it cannot eliminate it completely.

The specification of the set  $\mathcal{S}$  plays a crucial role in the worst-case approach and, in general, is a difficult task. Clearly selecting a wide set, increases the chances of including the real model  $\mathcal{P}$  and makes the risk measure more conservative; on the other hand if  $\mathcal{S}$  is too large  $\bar{\rho}$  might become unrealistic. Several choices have been considered in the literature and present different interpretations. In this contribution we assume  $\mathcal{S} = \{\mathcal{P}_1, \dots, \mathcal{P}_m\}$ , that is we consider a finite set of probability measures. This choice is rather frequent in a context of model ambiguity and it has also the benefit of making the problem mathematically tractable. A finite set of models is typical in situations where there is not enough evidence from data to select a model and the specification of  $\mathcal{S}$  is left to experts opinion. In the context of measuring market risk, the Standard portfolio Analysis of Risk (SPAN, 1995) proposed by the Chicago Mercantile Exchange provides an example of finite  $\mathcal{S}$  consisting of 16 probability measures obtained combining up and down movement of the volatility with up, down or no move of the future prices (see Artzner et al., 1999, Section 3.2 for a detailed description of how these scenarios are built). To reduce the impact of model risk in option pricing, Cont (2006) presents a worst-case approach over two probability measures: one providing a jump-diffusion model and the other one a simpler diffusion model, see Example 4.4 in his paper. In the insurance framework, an example of finite set  $\mathcal{S}$  is obtained considering the set of different Catastrophe Models provided by Cat modelling agencies; the insurer then has to take a robust decision with respect to these models, see for instance Calder, Couper, and Lo (2012).

A valid alternative for  $\mathcal{S}$  is to consider the convex hull of  $\{\mathcal{P}_1, \dots, \mathcal{P}_m\}$

$$\mathcal{S}' := \left\{ \mathcal{P} : \lambda \geq \mathbf{0}, \mathbf{1}^T \lambda = 1 \text{ and } \mathcal{P}(\cdot) = \sum_{k \in \mathcal{M}} \lambda_k \mathcal{P}_k(\cdot) \right\},$$

which is precisely what Zhu and Fukushima (2009) consider when  $\rho \equiv \text{CVaR}$ . It is shown that

$$\text{WCVaR}_\alpha(X) := \sup_{\mathcal{P} \in \mathcal{S}'} \rho(X; \mathcal{P}) = \min_{t \in \mathbb{R}} \max_{k \in \mathcal{M}} \left\{ t + \frac{1}{1 - \alpha} \mathbf{E}_{\mathcal{P}_k}(X - t)_+ \right\}, \tag{2.8}$$

where  $\mathbf{E}_{\mathcal{P}_k}(\cdot)$  is the expectation with respect to  $\mathcal{P}_k$ . Clearly,

$$\max_{k \in \mathcal{M}} \rho_{\mathcal{P}_k}(X) \leq \sup_{\mathcal{P} \in \mathcal{S}'} \rho(X; \mathcal{P}) \tag{2.9}$$

holds for any risk measure. Proposition 2.1 shows that the “worst-case” definitions are identical if  $\rho \equiv \text{VaR}$  and it is followed by an example showing that the above inequality may hold strictly if  $\rho \equiv \text{CVaR}$ .

**Proposition 2.1.** *Let  $\{\mathcal{P}_1, \dots, \mathcal{P}_m\}$  be a set of candidate models. Then,*

$$\max_{k \in \mathcal{M}} \text{VaR}_\alpha(X; \mathcal{P}_k) = \sup_{\mathcal{P} \in \mathcal{S}'} \text{VaR}_\alpha(X; \mathcal{P})$$

**Proof.** Without loss of generality, we may assume that  $m = 2$ . It is well-known that VaR has convex level sets, i.e. if two probability models  $\mathcal{P}_1, \mathcal{P}_2$  are such that  $\text{VaR}_\alpha(X; \mathcal{P}_1) = \text{VaR}_\alpha(X; \mathcal{P}_2)$ , then  $\text{VaR}_\alpha(X, \lambda \mathcal{P}_1 + (1 - \lambda) \mathcal{P}_2) = \text{VaR}_\alpha(X; \mathcal{P}_1)$  for any  $\lambda \in [0, 1]$  (see Gneiting, 2011). Further, VaR is monotone and translation invariant (see properties (a) and (b) from Section 2.3) and therefore, we can apply Lemma 2.2 in Bellini and Bignozzi (2015) to obtain that VaR is quasilinear. That is,

$$\begin{aligned} \min_{k \in \{1,2\}} \text{VaR}_\alpha(X; \mathcal{P}_k) &\leq \text{VaR}_\alpha(X; \lambda \mathcal{P}_1 + (1 - \lambda) \mathcal{P}_2) \\ &\leq \max_{k \in \{1,2\}} \text{VaR}_\alpha(X; \mathcal{P}_k). \end{aligned}$$

which in turn implies that

$$\sup_{\mathcal{P} \in \mathcal{S}} \text{VaR}_\alpha(X; \mathcal{P}) \leq \max_{k \in \{1,2\}} \text{VaR}_\alpha(X; \mathcal{P}_k).$$

The latter and (2.9) conclude the proof.  $\square$

As it has been anticipated, the same result does not hold for CVaR. Indeed, the following example shows that  $\max_{k \in \mathcal{M}} \text{CVaR}_\alpha(X; \mathcal{P}_k) < \sup_{\mathcal{P} \in \mathcal{S}} \text{CVaR}_\alpha(X; \mathcal{P})$  may hold.

**Example 2.1.** Consider a discrete random variable  $X$  which takes only four values, i.e.  $\{1, 2, 3, 4\}$ . We only consider two possible probability models,  $\mathcal{P}_1$  and  $\mathcal{P}_2$ , such that

$$\begin{aligned} \mathcal{P}_1(X = 1) &= 0, \mathcal{P}_1(X = 2) = \frac{1}{2}, \mathcal{P}_1(X = 3) = \frac{1}{6}, \mathcal{P}_1(X = 4) = \frac{1}{2}, \\ \mathcal{P}_2(X = 1) &= \frac{1}{2}, \mathcal{P}_2(X = 2) = 0, \mathcal{P}_2(X = 3) = 0, \mathcal{P}_2(X = 4) = \frac{1}{2}. \end{aligned}$$

It is not difficult to find that  $\text{CVaR}_{2/3}(X; \mathcal{P}_1) = \text{CVaR}_{2/3}(X; \mathcal{P}_2) = \frac{13}{4}$ . Let  $\mathcal{P}_0 = \frac{1}{2} \mathcal{P}_1 + \frac{1}{2} \mathcal{P}_2$ , i.e.

$$\mathcal{P}_0(X = 1) = \frac{1}{4}, \mathcal{P}_0(X = 2) = \frac{1}{4}, \mathcal{P}_0(X = 3) = \frac{1}{12}, \mathcal{P}_0(X = 4) = \frac{5}{12},$$

which is an element of  $\mathcal{S}'$ . Clearly,  $\text{CVaR}_{2/3}(X; \mathcal{P}_0) = \frac{27}{8}$ , which justifies our claim as follows:

$$\frac{13}{4} = \max_{k \in \{1,2\}} \text{CVaR}_\alpha(X; \mathcal{P}_k) < \text{CVaR}_\alpha(X; \mathcal{P}_0) \leq \sup_{\mathcal{P} \in \mathcal{S}'} \text{CVaR}_\alpha(X; \mathcal{P}).$$

2.3. Properties of the worst-case risk measure

Some properties of the worst-case risk measures have been briefly discussed in [Zhu and Fukushima \(2009\)](#) and therefore, we further outline the main traits of this class. We now restate some of the properties often satisfied by a risk measure and examine if its worst-case counterpart preserves these properties. Thus,

- (a) *Monotonicity*:  $\rho_{\mathcal{P}}(X) \leq \rho_{\mathcal{P}}(Y)$  holds if  $\mathcal{P}(X \leq Y) = 1$ .
- (b) *Translation Invariance*:  $\rho_{\mathcal{P}}(X - m) = \rho_{\mathcal{P}}(X) - m$  holds for any  $m \in \mathbb{R}$ .
- (c) *Positive homogeneity*:  $\rho_{\mathcal{P}}(\lambda X) = \lambda \rho_{\mathcal{P}}(X)$  holds for any  $\lambda > 0$ .
- (d) *Subadditivity*:  $\rho_{\mathcal{P}}(X + Y) \leq \rho_{\mathcal{P}}(X) + \rho_{\mathcal{P}}(Y)$ .
- (e) *Convexity*:  $\rho_{\mathcal{P}}(\beta X + (1 - \beta)Y) \leq \beta \rho_{\mathcal{P}}(X) + (1 - \beta)\rho_{\mathcal{P}}(Y)$  holds for any  $\beta \in (0, 1)$ .
- (f) *Comonotonic additivity*: If  $X, Y$  are comonotone, then  $\rho_{\mathcal{P}}(X + Y) = \rho_{\mathcal{P}}(X) + \rho_{\mathcal{P}}(Y)$ .
- (g) *Comonotonic subadditivity*: If  $X, Y$  are comonotone, then  $\rho_{\mathcal{P}}(X + Y) \leq \rho_{\mathcal{P}}(X) + \rho_{\mathcal{P}}(Y)$ .

where  $X, Y \in \mathcal{X}$ . By definition,  $X$  and  $Y$  are comonotone if

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geq 0 \quad \text{for any } \omega \times \omega' \in \Omega \times \Omega.$$

It is well-known that VaR satisfies properties (a)–(c) and (f)–(g), while CVaR fulfils all properties (a)–(g). Proposition 1 in [Zhu and Fukushima \(2009\)](#) also shows that  $\bar{\rho}(X) = \sup_{k \in \mathcal{M}} \rho_{\mathcal{P}_k}(X)$  satisfies (a)–(e) if  $\rho$  satisfies (a)–(e). Properties (f) and (g) are detailed in [Proposition 2.2](#).

**Proposition 2.2.** Let  $\{\mathcal{P}_1, \dots, \mathcal{P}_m\}$  be a set of candidate models and  $\rho$  be a risk measure that satisfies properties (f) (or (g)). Then,  $\bar{\rho}(X) = \sup_{k \in \mathcal{M}} \rho_{\mathcal{P}_k}(X)$  satisfies (g).

**Proof.** Let  $X$  and  $Y$  be comonotone and assume  $\rho$  is comonotonic additive, then

$$\begin{aligned} \bar{\rho}(X + Y) &= \sup_{k \in \mathcal{M}} \rho(X + Y; \mathcal{P}_k) = \sup_{k \in \mathcal{M}} \{\rho(X; \mathcal{P}_k) + \rho(Y; \mathcal{P}_k)\} \\ &\leq \bar{\rho}(X) + \bar{\rho}(Y). \end{aligned}$$

If  $\rho$  is comonotonic subadditive then it is sufficient to replace the second equality from above with a less than or equal to inequality.  $\square$

Relaxing the assumption of properties (f)–(g) is rather common in a model uncertainty setting, where no pre-specified reference probability measure is available (for example, see [Song & Yan, 2009](#)). The next example illustrates that  $\overline{\text{CVaR}}_{\alpha}(X) := \sup_{k \in \mathcal{M}} \text{CVaR}_{\alpha}(X; \mathcal{P}_k)$  may be strictly comonotonic subadditive, i.e.

$$\overline{\text{CVaR}}_{\alpha}(X + Y) < \overline{\text{CVaR}}_{\alpha}(X) + \overline{\text{CVaR}}_{\alpha}(Y).$$

**Example 2.2.** Consider a discrete random variable  $X$  which takes only three values, i.e.  $\{2, 3, 4\}$ . Let there be two comonotone random variables,  $X$  and  $X^2$ , and only two possible probability models  $\{\mathcal{P}_1, \mathcal{P}_2\}$  such that

$$\mathcal{P}_1(X=2) = \frac{3}{4}, \mathcal{P}_1(X=3) = \frac{1}{6}, \mathcal{P}_2(X=2) = 0.8, \mathcal{P}_2(X=3) = 0.08.$$

We compute CVaR at level  $\alpha = \frac{2}{3}$ . It is not difficult to find that

$$\text{CVaR}_{2/3}(X; \mathcal{P}_1) = 3, \text{CVaR}_{2/3}(X^2; \mathcal{P}_1) = 9.5,$$

$$\text{CVaR}_{2/3}(X + X^2; \mathcal{P}_1) = 12.5$$

$$\text{CVaR}_{2/3}(X; \mathcal{P}_2) = 2.96, \text{CVaR}_{2/3}(X^2; \mathcal{P}_2) = 9.52,$$

$$\text{CVaR}_{2/3}(X + X^2; \mathcal{P}_2) = 12.48,$$

and thus,  $\overline{\text{CVaR}}_{2/3}(X + X^2) < \overline{\text{CVaR}}_{2/3}(X) + \overline{\text{CVaR}}_{2/3}(X^2)$ , as previously claimed.

3. VaR robust optimisation

In this section, we solve the worst-case scenario optimisation problem (2.4) and worst-case regret optimisation problem (2.5) under the  $C^{co} \times \mathfrak{R}$  feasibility set, when the risk measure  $\rho$  is VaR. Note that

$$\text{VaR}_{\alpha}(X - R[X]; \mathcal{P}) = \text{VaR}_{\alpha}(X; \mathcal{P}) - R(\text{VaR}_{\alpha}(X; \mathcal{P})), \tag{3.1}$$

for any  $I \in C^{co}$ . For brevity, we denote  $a_k = \text{VaR}_{\alpha}(X; \mathcal{P}_k)$ .

3.1. Worst-case scenario VaR optimisation problem

We first observe that the objective function is increasing and continuous in  $P$  and any feasible premium  $P$  is bounded below by  $\omega_0 + (1 + \theta) \max_{k \in \mathcal{M}} \mathbb{H}_{\mathcal{P}_k}(R[X])$  for any fixed  $R \in C^{co}$ . The latter leads us to define the following subset of  $C^{co}$ , which essentially puts an upper bound on the set of feasible contracts:

$$C' = \left\{ R \in C^{co} \mid \omega_0 + (1 + \theta) \max_{k \in \mathcal{M}} \mathbb{H}_{\mathcal{P}_k}(R[X]) \leq \bar{P} \right\}.$$

Eq. (3.1) helps in justifying the next lemma.

**Lemma 3.1.** If Assumption 2.1 holds with  $\rho \equiv \text{VaR}_{\alpha}$ , then any contract  $R \in C^{co}$  is feasible for Problem 2.4 with a feasibility set  $C^{co} \times \mathfrak{R}$  if and only if  $R \in C'$ . Further, for any fixed  $R \in C'$ , the optimal premium is given by  $P_R^* = \omega_0 + (1 + \theta) \max_{k \in \mathcal{M}} \mathbb{H}_{\mathcal{P}_k}(R[X])$  and the optimisation problem from (2.4) is equivalent to

$$\min_{R \in C'} \left\{ P_R^* + \max_{k \in \mathcal{M}} (a_k - R[a_k]) \right\}. \tag{3.2}$$

Define now  $a^* = \max_{k \in \mathcal{M}} a_k$ . Since  $R \in C^{co}$ , the map  $x \mapsto x - R[x]$  is non-decreasing, and thus

$$\max_{k \in \mathcal{M}} (a_k - R[a_k]) = a^* - R[a^*] \quad \text{for all } R \in C.$$

Hence, Problem (3.2) becomes  $\min_{R \in C'} \{P_R^* + a^* - R[a^*]\}$ . By stratifying the set  $C'$  of feasible contracts according to the values  $\xi = R[a^*]$ , this optimisation problem can be decomposed into a two-step minimisation problem:

$$\min_{0 \leq \xi \leq a^*} \{a^* - \xi + \min_{R \in C'_{\xi}} P_R^*\}, \quad \text{where } C'_{\xi} = \{R \in C' \mid R[a^*] = \xi\}$$

$$\text{for any } 0 \leq \xi \leq a^*. \tag{3.3}$$

Due to the presence of the premium constraint  $P_R^* \leq \bar{P}$ , the set  $C'_{\xi}$  could be empty if  $\xi$  is too large. The next result explains the effective range of  $\xi$  of the outer minimisation of Problem (3.3). The proof relies on the simple observation that the insurance layer contract

$$R_{\xi}^*[X] = (X - a^* + \xi)_+ - (X - a^*)_+$$

belongs to  $C^{co}$  with  $R_{\xi}^*[a^*] = \xi$  and the fact that this contract is minimal in the following sense:

$$R_{\xi}^*[X] \leq R[X] \quad \text{for all } R \in C^{co} \text{ with } R[a^*] = \xi.$$

**Lemma 3.2.** If Assumption 2.1 holds with  $\rho \equiv \text{VaR}_{\alpha}$ , then for any  $\xi \in [0, a^*]$ , the set  $C'_{\xi}$  is non-empty if and only if

$$0 \leq \xi \leq \Delta = \max \left\{ \xi \leq a^* \mid \omega_0 + (1 + \theta) \max_{k \in \mathcal{M}} \mathbb{H}_{\mathcal{P}_k}(R_{\xi}^*[X]) \leq \bar{P} \right\}.$$

**Proof.** If  $0 \leq \xi \leq \Delta$ , the contract  $R_{\xi}^*$  belongs to  $C'_{\xi}$  by construction. To prove the converse, suppose that there exists a contract  $R \in C'_{\xi}$  with  $\xi > \Delta$ . Since  $R[X] \geq R_{\xi}^*[X]$ , we have

$$\omega_0 + (1 + \theta) \max_{k \in \mathcal{M}} \mathbb{H}_{\mathcal{P}_k}(R[X]) \geq \omega_0 + (1 + \theta) \max_{k \in \mathcal{M}} \mathbb{H}_{\mathcal{P}_k}(R_{\xi}^*[X]) > \bar{P},$$

which contradicts the definition of  $C'$ .  $\square$

Recall that the premium principle  $\mathbb{H}$  is monotone. By Lemma 3.2 and the minimality of the insurance layer contract  $R_\xi^*$  in the set  $\mathcal{C}'_\xi$  for  $0 \leq \xi \leq \Delta$ , the inner minimisation of Problem (3.3) is solved by the contract  $R_\xi^*$ , whenever  $0 \leq \xi \leq \Delta$ . Therefore, it remains to obtain the optimal value of  $\xi$  for the outer minimisation, which is essentially a one-dimensional problem. We summarise our findings for the worst-case scenario VaR optimisation problem in the next theorem.

**Theorem 3.1.** *If Assumption 2.1 holds with  $\rho \equiv \text{VaR}_\alpha$ , then the solution  $(R^*, P^*)$  of Problem (2.4), assumed to be solved over the set  $\mathcal{C}^{\text{co}} \times \mathfrak{N}$ , is given by*

$$R^*[X] = R_{\xi^*}^*[X] = (X - a^* + \xi^*)_+ - (X - a^*)_+$$

$$\text{and } P^* = \omega_0 + (1 + \theta) \max_{k \in \mathcal{M}} \mathbb{H}_{\mathcal{P}_k} (R^*[X]),$$

where  $\xi^*$  is a solution of

$$\min_{0 \leq \xi \leq \Delta} \left\{ a^* - \xi + \omega_0 + (1 + \theta) \max_{k \in \mathcal{M}} \mathbb{H}_{\mathcal{P}_k} (R_\xi^*[X]) \right\}. \tag{3.4}$$

Moreover, the optimal objective value is  $a^* - \xi^* + P^*$ .

**Remark 3.1.** The solution of (3.4) is unique as long as  $g$  and all  $F_i$  are strictly increasing functions.

In the rest of this section, we demonstrate how Problem (3.4) can be solved rather explicitly, whenever  $\mathbb{H}$  is a distortion premium principle as given in (2.1). Note that  $g$  is non-decreasing, and thus,  $H$  holds the comonotonic additivity property (for details, see Dhaene, Kukush, Linders, & Tang, 2012). Consequently,

$$\begin{aligned} \mathbb{H}_{\mathcal{P}_k} (R_\xi^*[X]) &= \mathbb{H}_{\mathcal{P}_k} ((X - a^* + \xi)_+) - \mathbb{H}_{\mathcal{P}_k} ((X - a^*)_+) \\ &= \int_{a^* - \xi}^{a^*} g(\bar{F}_k(t)) dt. \end{aligned}$$

Also, the above function is convex in  $\xi \in [0, a^*]$  for any  $k \in \mathcal{M}$ , since  $g$  is non-decreasing. Thus,

$$G(\xi) = a^* - \xi + \omega_0 + (1 + \theta) \max_{k \in \mathcal{M}} \mathbb{H}_{\mathcal{P}_k} (R_\xi^*[X])$$

is convex in  $\xi \in [0, \Delta]$ . Therefore, Problem (3.4) can be solved by finding the directional derivatives of  $G$ . To this end, we define the directional derivative of an arbitrary convex function  $H$  at  $\xi$  along the direction  $d \in \mathfrak{N}$ , if exists, as

$$H'(\xi; d) = \lim_{t \searrow 0} \frac{H(\xi + td) - H(\xi)}{t},$$

which is positively homogeneous in  $d \in \mathfrak{N}$ . The right-hand and left-hand derivatives of  $H$  can be expressed as  $H'_+(\xi) = H'(\xi; 1)$  and  $H'_-(\xi) = -H'(\xi; -1)$ , respectively. A point  $\xi^* \in [0, \Delta]$  is optimal for Problem (3.4) if and only if it satisfies the first order condition  $G'(\xi^*; \xi - \xi^*) \geq 0$  for all  $\xi \in [0, \Delta]$ . The next lemma provides a simplified expression for the directional derivative of  $G$ .

**Lemma 3.3.** *Assume that  $\mathbb{H}$  satisfies (2.1) and denote*

$$A(\xi) = \left\{ i \in \mathcal{M} \mid \mathbb{H}_{\mathcal{P}_i} (R_\xi^*[X]) = \max_{k \in \mathcal{M}} \mathbb{H}_{\mathcal{P}_k} (R_\xi^*[X]) \right\},$$

for any  $0 \leq \xi \leq \Delta$ . The directional derivative of  $G$  at  $\xi$  along the direction  $d \in \mathfrak{N}$  is given by

$$G'(\xi; d) = \begin{cases} d\{-1 + (1 + \theta) \max_{i \in A(\xi)} g(\bar{F}_i((a^* - \xi)_-))\}, & d \geq 0, \\ d\{-1 + (1 + \theta) \min_{i \in A(\xi)} g(\bar{F}_i(a^* - \xi))\}, & d < 0. \end{cases}$$

**Proof.** For each  $k \in \mathcal{M}$ , let

$$g_k(\xi) = \int_{a^* - \xi}^{\infty} g(\bar{F}_k(t)) dt - \mathbb{H}_{\mathcal{P}_k} ((X - a^*)_+), \quad \xi \in [0, \Delta].$$

Its right-hand and left-hand derivatives at  $\xi$  are given by

$$g'_k(\xi, 1) = g(\bar{F}_k((a^* - \xi)_-)) \quad \text{and} \quad -g'_k(\xi, -1) = g(\bar{F}_k(a^* - \xi)),$$

respectively. Therefore, the directional derivative of  $g_k$  at  $\xi \in [0, \Delta]$  along the direction  $d \in \mathfrak{N}$  equals to

$$g'_k(\xi; d) = \begin{cases} d g(\bar{F}_k((a^* - \xi)_-)), & d \geq 0, \\ d g(\bar{F}_k(a^* - \xi)), & d < 0. \end{cases}$$

Our claim follows from the classical Danskin's Theorem (see for example, Corollary 1.30 of Güler, 2010), which asserts that

$$(\max_{k \in \mathcal{M}} g_k)'(\xi; d) = \max_{i \in A(\xi)} g'_i(\xi; d), \quad d \in \mathfrak{N}.$$

The proof is now complete.  $\square$

### 3.2. Worst-case regret VaR optimisation problem

We turn our attention to the worst-case regret VaR optimisation from (2.5). Since we are no longer able to use similar argumentation as in the previous subsection, the usual approach in the existing literature is to assume a discrete distributed  $X$ . That is,  $X = \{x_1, \dots, x_n\}$ , where without loss of generality, it can be assumed that  $x_1 \leq \dots \leq x_n$ . Let us denote  $p_{ik} = \mathcal{P}_k(X = x_i)$ . Clearly,  $\mathbf{p}_k \geq \mathbf{0}$  and  $\mathbf{1}^T \mathbf{p}_k = 1$  for all  $k \in \mathcal{M}$ , where  $\mathbf{0}$  and  $\mathbf{1}$  are the  $n$ -dimensional column vector of zeroes and ones, respectively. By convention, the inequality and equality between two vectors is understood componentwise. Denote  $R[x_i] = y_i$  and if  $R \in \mathcal{C}^{\text{co}}$ , then we should have

$$0 \leq y_i \leq x_i \text{ and } 0 \leq y_i - y_{i-1} \leq x_i - x_{i-1}, \text{ for all } i \in \mathcal{M},$$

where by convention  $y_0 = x_0 = 0$ . The above can be rewritten as  $\mathbf{0} \leq \mathbf{y} \leq \mathbf{x}$  and  $\mathbf{0} \leq \mathbf{A}\mathbf{y} \leq \mathbf{A}\mathbf{x}$  with the  $(n - 1) \times n$  matrix  $\mathbf{A}$  given by

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & \dots & 0 & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & \dots & -1 & 1 \end{pmatrix}.$$

Since  $\mathbf{x} - \mathbf{y}$  is increasingly ordered, then

$$\text{VaR}_\alpha(X - R[X]) = x_{p(k)} - y_{p(k)}, \text{ where } p(k) = \min_j \left\{ \sum_{i=1}^j p_{ik} \geq \alpha \right\}.$$

In order to make our optimisation problems tractable, we assume that  $\mathbb{H}$  satisfies (2.1). Thus,

$$\mathbb{H}_{\mathcal{P}_k} [X] = \boldsymbol{\pi}_k^T \mathbf{x} \tag{3.5}$$

as a result of Dhaene et al. (2012). Specifically, If  $g$  is a left continuous function, then

$$\pi_{ik} = g\left(1 - \sum_{j=1}^{i-1} p_{jk}\right) - g\left(1 - \sum_{j=1}^i p_{jk}\right), \quad 1 \leq i \leq n, \quad k \in \mathcal{M}.$$

Consequently, Problem (2.5) is an LP and we state this result as Proposition 3.1.

**Proposition 3.1.** *Let Assumption 2.1 hold with  $\rho \equiv \text{VaR}_\alpha$ . If  $X$  is a discrete random variable that takes the values  $\{x_1, \dots, x_n\}$  such that  $x_1 \leq \dots \leq x_n$  and  $\mathbb{H}$  satisfies (2.1), then solving Problem (2.5) over the set  $\mathcal{C}^{\text{co}} \times \mathfrak{N}$  is equivalent to solving*

$$\begin{cases} \min_{(\mathbf{y}, P, r) \in \mathfrak{N}^n \times \mathfrak{N} \times \mathfrak{N}} & r, \\ \text{s.t.} & x_{\pi(k)} - y_{\pi(k)} + P - \rho_k^* \leq r, \quad k \in \mathcal{M}, \\ & \omega_0 + (1 + \theta) \boldsymbol{\pi}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ & \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}, \\ & \mathbf{0} \leq \mathbf{A}\mathbf{y} \leq \mathbf{A}\mathbf{x}. \end{cases}$$

where  $\rho_k^*$  is the optimal objective value of Problem (2.6). That is,

$$\left\{ \begin{array}{l} \rho_k^* = \min_{(\mathbf{y}, P) \in \mathbb{R}^n \times \mathbb{R}} \{x_{p(k)} - y_{p(k)} + P\}, \\ \text{s.t. } \omega_0 + (1 + \theta)\boldsymbol{\pi}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}, \\ \mathbf{0} \leq \mathbf{A}\mathbf{y} \leq \mathbf{A}\mathbf{x}. \end{array} \right.$$

**Remark 3.2.** Keeping the same set of assumptions as given in Proposition 3.1, solving Problem (2.4) over the set  $C^{co} \times \mathbb{R}$  is equivalent to solving

$$\left\{ \begin{array}{l} \min_{(\mathbf{y}, P, r) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}} r, \\ \text{s.t. } x_{\pi(k)} - y_{\pi(k)} + P \leq r, \quad k \in \mathcal{M}, \\ \omega_0 + (1 + \theta)\boldsymbol{\pi}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}, \\ \mathbf{0} \leq \mathbf{A}\mathbf{y} \leq \mathbf{A}\mathbf{x}. \end{array} \right.$$

**Remark 3.3.** Due to relation (3.5), a variant of the LP reformulations from Proposition 3.1 and Remark 3.2 can be written for any case in which the risk measure  $\rho$  is a distortion risk measure. The key assumption is that  $R \in C^{co}$  and the fact that distortion risk measures are comonotonic and thus, Problems (2.4)–(2.6) can be reformulated as LPs for any comonotone additive risk measure  $\rho$ . For example, any risk measure that satisfies (2.1) is comonotone additive (see Dhaene et al., 2012). Note that  $\mathbf{y}$  and  $\mathbf{x} - \mathbf{y}$  are increasingly ordered (as  $\mathbf{x}$  is increasingly ordered), which make the optimisation problems under the set  $C^{co} \times \mathbb{R}$  tractable. The lack of ordering could be overcome only if  $\rho \equiv \text{CVaR}_\alpha$ , as it can be seen in Section 4, where the comonotonic assumption is removed. Finally, if the cost of insurance follows a different premium calculation, i.e. does not satisfy (2.1), then the corresponding constraints may not be linear, but are Second-order cone programming (SOCP) representable for any well-known premium calculations (for details, see Asimit, Gao, Hu, & Kim, 2017), case in which, we only require  $I \in C^{co}$ , i.e.  $\mathbf{x} - \mathbf{y}$  is increasingly ordered, in order to preserve the linearity of the objective functions. Thus, if  $\mathbb{H}$  does not satisfy (2.1), the optimisation problems are of SOCP-type.

#### 4. CVaR robust optimisation

The current section provides numerical solutions to the CVaR-type of Problems (2.4) and (2.5) under similar assumptions to the ones made in Section 3.2. The crucial change is made by the fact that the set of feasible solutions, namely  $C \times \mathbb{R}$ , is larger and moral hazard is permitted. Recall that if moral hazard is excluded, then the optimisation problems could have been solved as in Section 3.2 (for details, see Remark 3.3). Moreover, the rationality constraints,  $0 \leq R[X] \leq X$ , are still required. With the help of Eq. (2.3), Problem (2.4) can be rewritten as follows:

$$\left\{ \begin{array}{l} \min_{(\mathbf{y}, P) \in \mathbb{R}^n \times \mathbb{R}} \max_{k \in \mathcal{M}} \min_{t \in \mathbb{R}^m} \left\{ t_k + \frac{1}{1-\alpha} \mathbf{p}_k^T (\mathbf{x} - \mathbf{y} - \mathbf{1}t_k)_+ + P \right\}, \\ \text{s.t. } \omega_0 + (1 + \theta)\mathbf{p}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}. \end{array} \right. \quad (4.1)$$

where without loss of generality  $\mathbb{H}_{\mathcal{P}_k} = \mathbb{E}_{\mathcal{P}_k}$  with  $k \in \mathcal{M}$  could be assumed (see Remark 3.3). In addition, Problem (2.5) is given by

$$\left\{ \begin{array}{l} \min_{(\mathbf{y}, P) \in \mathbb{R}^n \times \mathbb{R}} \max_{k \in \mathcal{M}} \min_{t \in \mathbb{R}^m} \left\{ t_k + \frac{1}{1-\alpha} \mathbf{p}_k^T (\mathbf{x} - \mathbf{y} - \mathbf{1}t_k)_+ + P - \rho_k^* \right\}, \\ \text{s.t. } \omega_0 + (1 + \theta)\mathbf{p}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}. \end{array} \right. \quad (4.2)$$

Recall that  $\rho_k^*$  represents the optimal value of the objective function from Problem (2.6) with  $\rho \equiv \text{CVaR}_\alpha$  and is given by

$$\left\{ \begin{array}{l} \rho_k^* = \min_{(t, \mathbf{y}, P) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}} \left\{ t + \frac{1}{1-\alpha} \mathbf{p}_k^T (\mathbf{x} - \mathbf{y} - \mathbf{1}t)_+ + P \right\}, \\ \text{s.t. } \omega_0 + (1 + \theta)\boldsymbol{\pi}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}. \end{array} \right. \quad (4.3)$$

In the remaining part of this section, we show that Problems (4.1)–(4.3) can be reduced to LP reformulations. The next theorem deals with Problem (4.1).

**Theorem 4.1.** Let Assumption 2.1 hold with  $\rho \equiv \text{CVaR}_\alpha$ . If  $X$  is a discrete random variable that takes the values  $\{x_1, \dots, x_n\}$  and  $\mathbb{H}_{\mathcal{P}_k} = \mathbb{E}_{\mathcal{P}_k}$  for all  $k \in \mathcal{M}$ , then solving Problem (4.1) over the set  $C \times \mathbb{R}$  is equivalent to

$$\left\{ \begin{array}{l} \min_{(\mathbf{t}, \mathbf{y}, \boldsymbol{\xi}, P, Z) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}} Z, \\ \text{s.t. } t_k + \frac{1}{1-\alpha} \mathbf{p}_k^T \boldsymbol{\xi}_k + P \leq Z, \quad k \in \mathcal{M}, \\ \mathbf{x} - \mathbf{y} - \mathbf{1}t_k \leq \boldsymbol{\xi}_k, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \boldsymbol{\xi}_k, \quad k \in \mathcal{M}, \\ \omega_0 + (1 + \theta)\mathbf{p}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}. \end{array} \right. \quad (4.4)$$

**Proof.** Let  $\boldsymbol{\xi}_k = (\xi_{1k}, \dots, \xi_{nk})^T$  for all  $k \in \mathcal{M}$ . Then, (4.1) may be equivalently formulated as:

$$\left\{ \begin{array}{l} \min_{(\mathbf{y}, \boldsymbol{\xi}, P) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}} \max_{k \in \mathcal{M}} \min_{t \in \mathbb{R}^m} \left\{ t_k + \frac{1}{1-\alpha} \mathbf{p}_k^T \boldsymbol{\xi}_k + P \right\}, \\ \text{s.t. } \mathbf{x} - \mathbf{y} - \mathbf{1}t_k \leq \boldsymbol{\xi}_k, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \boldsymbol{\xi}_k, \quad k \in \mathcal{M}, \\ \omega_0 + (1 + \theta)\mathbf{p}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}. \end{array} \right. \quad (4.5)$$

Note that the objective function in (4.5) is increasing in  $\xi_{ik}$  for all  $1 \leq i \leq n$  and  $k \in \mathcal{M}$ . Thus, the first two constraints from the latter optimisation problem ensure that  $\boldsymbol{\xi}_k = (\mathbf{x} - \mathbf{y} - \mathbf{1}t_k)_+$  for  $k \in \mathcal{M}$ . Thus, (4.5) can be rewritten as follows:

$$\left\{ \begin{array}{l} \min_{(\mathbf{y}, \boldsymbol{\xi}, P, Z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}} Z, \\ \text{s.t. } \min_{t \in \mathbb{R}^m} \left\{ t_k + \frac{1}{1-\alpha} \mathbf{p}_k^T \boldsymbol{\xi}_k + P \right\} \leq Z, \quad k \in \mathcal{M}, \\ \mathbf{x} - \mathbf{y} - \mathbf{1}t_k \leq \boldsymbol{\xi}_k, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \boldsymbol{\xi}_k, \quad k \in \mathcal{M}, \\ \omega_0 + (1 + \theta)\mathbf{p}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}. \end{array} \right. \quad (4.6)$$

We now show that  $(\mathbf{y}^*, \boldsymbol{\xi}^*, P^*, z^*)$  solves Problem (4.6) if and only if  $(\mathbf{t}^*, \mathbf{y}^*, \boldsymbol{\xi}^*, P^*, z^*)$  solves Problem (4.4), where

$$\mathbf{t}^* = \operatorname{argmin}_{t \in \mathbb{R}^m} \left\{ t_k + \frac{1}{1-\alpha} \mathbf{p}_k^T \boldsymbol{\xi}_k^* + P^* \right\} \text{ for all } k \in \mathcal{M}. \quad (4.7)$$

Suppose that  $(\mathbf{y}^*, \boldsymbol{\xi}^*, P^*, z^*)$  solves Problem (4.6), which implies that  $(\mathbf{t}^*, \mathbf{y}^*, \boldsymbol{\xi}^*, P^*, z^*)$  is a feasible solution to Problem (4.4). If  $(\mathbf{t}^*, \mathbf{y}^*, \boldsymbol{\xi}^*, P^*, z^*)$  does not solve Problem (4.4), then there exists a feasible solution  $(\mathbf{t}', \mathbf{y}', \boldsymbol{\xi}', P', z')$  such that  $z' < z^*$ . Now, for all  $k \in \mathcal{M}$  we have that

$$\min_{t \in \mathbb{R}^m} \left\{ t_k + \frac{1}{1-\alpha} \mathbf{p}_k^T \boldsymbol{\xi}_k + P \right\} \leq t'_k + \frac{1}{1-\alpha} \mathbf{p}_k^T \boldsymbol{\xi}'_k + P' \leq z'. \quad (4.8)$$

Thus,  $(\mathbf{y}', \boldsymbol{\xi}', P', z')$  is feasible to Problem (4.6), which contradicts that  $(\mathbf{y}^*, \boldsymbol{\xi}^*, P^*, z^*)$  is an optimal solution to Problem (4.6).

Conversely, suppose that  $(\mathbf{t}', \mathbf{y}', \xi', P', z')$  solves Problem (4.4). Eq. (4.8) implies that  $(\mathbf{y}', \xi', P', z')$  is feasible to Problem (4.6). If  $(\mathbf{y}', \xi', P', z')$  does not solve Problem (4.6), then there exists a feasible solution  $(\mathbf{y}^*, \xi^*, P^*, z^*)$  such that  $z^* < z'$ . Then,  $(\mathbf{t}^*, \mathbf{y}^*, \xi^*, P^*, z^*)$  solves Problem (4.4), where  $\mathbf{t}^*$  is defined as in (4.7). The latter contradicts our initial assumption that  $(\mathbf{t}', \mathbf{y}', \xi', P', z')$  is an optimal solution to Problem (4.4). The proof is now complete.  $\square$

Finally, we solve Problems (4.2) and (4.3). By following the same arguments as provided in the proof of Theorem 4.1, one may show our claims from Proposition 4.1, and therefore, the proofs are left to the reader.

**Proposition 4.1.** *Let Assumption 2.1 hold with  $\rho \equiv \text{CVaR}_\alpha$ . If  $X$  is a discrete random variable that takes the values  $\{x_1, \dots, x_n\}$  and  $\mathbb{E}_{P_k} = \mathbb{E}_{P_k}$  for all  $k \in \mathcal{M}$ , then solving Problem (4.3) over the set  $\mathcal{C} \times \mathfrak{N}$  is equivalent to*

$$\left\{ \begin{array}{l} \rho_k^* = \min_{(\mathbf{t}, \mathbf{y}, \xi, P) \in \mathfrak{N} \times \mathfrak{N}^m \times \mathfrak{N}^m \times \mathfrak{N}} \left\{ t + \frac{1}{1-\alpha} \mathbf{p}_k^T \xi + P \right\}, \\ \text{s.t.} \quad \mathbf{x} - \mathbf{y} - \mathbf{1}t \leq \xi, \\ \mathbf{0} \leq \xi, \\ \omega_0 + (1 + \theta) \mathbf{p}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}. \end{array} \right.$$

Moreover, Problem (4.2) is equivalent to

$$\left\{ \begin{array}{l} \min_{(\mathbf{t}, \mathbf{y}, \xi, P, z) \in \mathfrak{N} \times \mathfrak{N}^m \times \mathfrak{N}^m \times \mathfrak{N} \times \mathfrak{N}} z, \\ \text{s.t.} \quad t_k + \frac{1}{1-\alpha} \mathbf{p}_k^T \xi_k + P - \rho_k^* \leq z, \quad k \in \mathcal{M}, \\ \mathbf{x} - \mathbf{y} - \mathbf{1}t_k \leq \xi_k, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \xi_k, \quad k \in \mathcal{M}, \\ \omega_0 + (1 + \theta) \mathbf{p}_k^T \mathbf{y} \leq P \leq \bar{P}, \quad k \in \mathcal{M}, \\ \mathbf{0} \leq \mathbf{y} \leq \mathbf{x}. \end{array} \right.$$

**5. Pareto robust optimisation**

Robust optimal contracts have been found in Sections 3 and 4 without discussing the drawbacks and possible remedies of our proposed robust solutions. One major issue is when there are multiple robust solutions and we explain our point by considering the following general worst-case optimisation problem:

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{k \in \mathcal{M}} f_k(\mathbf{x}), \quad \text{with } f_k : \mathfrak{N}^n \rightarrow \mathfrak{N}, \tag{5.1}$$

where  $\mathcal{X} \in \mathfrak{N}^n$  is a non-empty feasibility set. Denote  $\mathcal{X}^{\text{Ro}} = \text{argmin}_{\mathbf{x} \in \mathcal{X}} \max_{k \in \mathcal{M}} f_k(\mathbf{x})$  the Robust solution set corresponding to (5.1). Identifying the Pareto solutions is a classical problem in economics, since those solutions make the allocation amongst various players as fair as possible, in the sense that no improvement could be made for one or more players without affecting the allocation of at least one player. The mathematical formulation of the Pareto solution set corresponding to (5.1) is given by:

$$\mathcal{X}^{\text{Pa}} = \{ \mathbf{x} \in \mathcal{X} \mid \nexists \tilde{\mathbf{x}} \in \mathcal{X} \text{ s.t. } f_k(\mathbf{x}) \geq f_k(\tilde{\mathbf{x}})$$

for all  $k \in \mathcal{M}$  and at least one inequality is strict  $\}$ .

It is not surprising that a Pareto solution may not be an element of  $\mathcal{X}^{\text{Ro}}$ , since worst-case type solutions are concerned only with extreme scenarios. Further,  $\mathbf{x}^* \in \mathcal{X}^{\text{Ro}}$  does not always imply that  $\mathbf{x}^* \in \mathcal{X}^{\text{Pa}}$ , when (5.1) admits multiple solutions. It is not difficult to show that if  $\mathbf{x}^*$  is the unique solution of (5.1) then,  $\mathbf{x}^* \in \mathcal{X}^{\text{Pa}}$ . Therefore, it is possible to solve (5.1) and produce a robust solution that is suboptimal for all concurrent objectives, which plays havoc with the entire decision process. Recall that Remark 3.1 explains when the closed-form solution is unique and one may show in that case that the unique solution is Pareto optimal as well.

Appa (2002) and Mangasarian (1979) provide methodologies to check the uniqueness property of an LP and therefore, there is no

issue with linear-type (5.1) optimisation problems with a unique solution. It is still not clear how to verify if a solution of (5.1) is an element of  $\mathcal{X}^{\text{Pa}}$ . In addition, it would be interesting to provide a constructive method to generate solutions from  $\mathcal{X}^{\text{Ro}} \cap \mathcal{X}^{\text{Pa}}$ . These are the aims of this section. Specifically, we first note that the discrete versions of (2.4) and (2.5) have the following linear representation:

$$\min_{\mathbf{x} \in \mathfrak{N}^n} \max_{k \in \mathcal{M}} \mathbf{c}_k^T \mathbf{x} + d_k, \quad \text{s.t. } \mathbf{A}_k \mathbf{x} \leq \mathbf{b}_k, \quad k \in \mathcal{M}, \tag{5.2}$$

with known  $\mathbf{A}_k$ ,  $\mathbf{b}_k$ ,  $\mathbf{c}_k$ ,  $\mathbf{d}_k$  matrices and column vectors of appropriate dimensions and known scalars  $d_k$ . The main result of this section, stated as Theorem 5.1, simply says how to always find a Pareto and robust optimal solution for (5.2) by solving at most one additional LP. These results are inspired by Theorem 1 of Iancu and Trichakis (2014) that solves a similar linear problem, where the decision-maker perceives the uncertainty in a very different way.

**Theorem 5.1.** *Let  $\mathbf{x}^*$  be any optimal solution of (5.2), where the latter problem is assumed to be non-trivial, i.e.  $\mathbf{c}_k$  with  $k \in \mathcal{M}$  are not all null vectors. Consider the following optimisation problem:*

$$\min_{\mathbf{y}} \sum_{k \in \mathcal{M}} \mathbf{c}_k^T \mathbf{y}, \quad \text{s.t. } \mathbf{A}_k(\mathbf{x}^* + \mathbf{y}) \leq \mathbf{b}_k, \quad \mathbf{c}_k^T \mathbf{y} \leq 0, \quad k \in \mathcal{M}. \tag{5.3}$$

If the optimal value in (5.3) is zero, then  $\mathbf{x}^* \in \mathcal{X}^{\text{Ro}} \cap \mathcal{X}^{\text{Pa}}$  in (5.2). If the optimal value in (5.3) is negative, then  $\mathbf{x}^* + \mathbf{y}^* \in \mathcal{X}^{\text{Ro}} \cap \mathcal{X}^{\text{Pa}}$  in (5.2), where  $\mathbf{y}^*$  is an optimal solution of (5.3).

**Proof.** It is not difficult to find that the objective function of (5.3) is always non-positive. Assume now that the objective function is zero such that  $\mathbf{x}^*$  is not Pareto solution, but is a robust optimal solution of (5.2). Therefore, there exists a feasible solution  $\hat{\mathbf{x}}$  of (5.2) such that

$$\mathbf{c}_k^T \mathbf{x}^* + d_k \geq \mathbf{c}_k^T \hat{\mathbf{x}} + d_k, \quad \text{for all } k \in \mathcal{M} \tag{5.4}$$

and at least one inequality holds strictly. Denote  $\hat{\mathbf{y}} = \hat{\mathbf{x}} - \mathbf{x}^*$  and since  $\hat{\mathbf{x}}$  is a feasible solution of (5.2), we get that

$$\mathbf{A}_k(\mathbf{x}^* + \hat{\mathbf{y}}) = \mathbf{A}_k \hat{\mathbf{x}} \leq \mathbf{b}_k, \quad \text{for all } k \in \mathcal{M}.$$

Recall that Eq. (5.4) tells us that  $\mathbf{c}_k^T \hat{\mathbf{x}} \leq \mathbf{c}_k^T \mathbf{x}^*$  for all  $k \in \mathcal{M}$ , and in turn we get that

$$\mathbf{c}_k^T \hat{\mathbf{y}} = \mathbf{c}_k^T \hat{\mathbf{x}} - \mathbf{c}_k^T \mathbf{x}^* \leq 0, \quad \text{for all } k \in \mathcal{M}.$$

Thus,  $\hat{\mathbf{y}}$  is a feasible solution of (5.3). Moreover, Eq. (5.4) suggests that  $\mathbf{c}_k^T \hat{\mathbf{x}} < \mathbf{c}_k^T \mathbf{x}^*$  for some  $k \in \mathcal{M}$ , and therefore,  $\mathbf{c}_k^T \hat{\mathbf{y}}$  is negative for some  $k \in \mathcal{M}$ . Consequently, the optimal objective value in (5.3) is negative, which contradicts our assumption of a null optimal objective value.

Assume now that the optimal objective value in (5.3) is negative. Note first that  $\mathbf{x}^* + \mathbf{y}^*$  is feasible in (5.2), since it is feasible in (5.3). Assume that  $\mathbf{x}^* + \mathbf{y}^*$  is not a Pareto solution, but is a robust optimal solution of (5.2). Thus, there exists a feasible solution,  $\tilde{\mathbf{x}}$ , of (5.2) such that  $\mathbf{x}^* + \mathbf{y}^*$  is Pareto dominated by  $\tilde{\mathbf{x}}$ . The mathematical formulation of the former is that

$$\mathbf{c}_k^T(\mathbf{x}^* + \mathbf{y}^*) + d_k \geq \mathbf{c}_k^T \tilde{\mathbf{x}} + d_k, \quad \text{for all } k \in \mathcal{M} \tag{5.5}$$

and at least one inequality holds strictly. Denote  $\tilde{\mathbf{y}} = \tilde{\mathbf{x}} - \mathbf{x}^*$  and since  $\tilde{\mathbf{x}}$  is a feasible solution of (5.2), one may find that  $\mathbf{x}^* + \tilde{\mathbf{y}}$  is a feasible solution to (5.2) as follows:

$$\mathbf{A}_k(\mathbf{x}^* + \tilde{\mathbf{y}}) = \mathbf{A}_k \tilde{\mathbf{x}} \leq \mathbf{b}_k, \quad \text{for all } k \in \mathcal{M}.$$

Now, Eq. (5.5) and the fact that  $\mathbf{y}^*$  is feasible solution for (5.3) imply that

$$\mathbf{c}_k^T \tilde{\mathbf{y}} = \mathbf{c}_k^T \tilde{\mathbf{x}} - \mathbf{c}_k^T \mathbf{x}^* \leq \mathbf{c}_k^T \mathbf{y}^* \leq 0,$$

which shows that  $\tilde{\mathbf{y}}$  is feasible in (5.3). We also know that at least one inequality from Eq. (5.5) is a strict inequality and as a result, one of the above inequality holds strictly, which results in  $\sum_{k \in \mathcal{M}} \mathbf{c}_k^T \tilde{\mathbf{y}} < \sum_{k \in \mathcal{M}} \mathbf{c}_k^T \mathbf{y}^*$ . The latter contradicts that  $\mathbf{y}^*$  is an optimal solution of (5.3). The proof is now complete.  $\square$



### 6. Numerical analysis

This section provides numerical illustrations to our worst-case scenario and regret optimisation problems from (2.4) and (2.5), respectively. Recall that in order to empirically solve these problems, a sample  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ , is drawn from the underlying distribution of  $X$ , and in turn, we find the optimal insurance contract  $\mathbf{y}^* = (y_1^*, y_2^*, \dots, y_n^*)^T$  and the optimal premium  $P^*$ . Let  $(\mathbf{y}_{wc}^*, P_{wc}^*)$  and  $(\mathbf{y}_{wr}^*, P_{wr}^*)$  denote the empirical optimal solutions to our robust models (2.4) and (2.5), respectively. Our main aim is to give a quality comparison between  $(\mathbf{y}_w^*, P_w^*)$ ,  $w \in \{wc, wr\}$ , and a best possible choice  $(\mathbf{y}^*, P^*)$ . Essentially, the latter is the “best solution” based on estimating a particular model chosen via two well-known standard statistical goodness-of-fit methods, namely *Akaike Information Criterion* (AIC) and *Corrected Akaike Information Criterion* (AICc), which are denoted as  $(\mathbf{y}_{AIC}^*, P_{AIC}^*)$  and  $(\mathbf{y}_{AICc}^*, P_{AICc}^*)$ , respectively. We believe that those comparisons are fair and explain the advantages and disadvantages of robust optimisation over a standard optimisation after choosing the most significant model (in the statistical sense). Finally, recall that all optimisations are implemented on a desktop with 6 core Intel i7-5820K at 3.30 gigahertz, 16 gigabytes RAM, running Linux x64, MATLAB R2014b, CVX 2.1.

The parameterisation employed in our empirical optimisation assumes that the loss variable  $X$  is LogNormal distributed with mean  $E(X) = 5,000$  and standard deviation  $\sqrt{3} \times E(X)$ . An expected value premium principle with a risk loading factor  $\theta = 0.25$  and no fixed/administrative costs, i.e.  $\omega_0 = 0$ , is assumed. In addition,  $\bar{p} = \frac{(1+\theta)E(X)}{2}$ . Since the underlying loss distribution of  $X$  is unknown, five candidate models are further assumed by the decision-maker:

- (i) *Model 1*: Exponential distribution with mean  $1/\nu$ .
- (ii) *Model 2*: LogNormal distribution with parameters  $(\mu, \sigma^2)$ .
- (iii) *Model 3*: Pareto distribution with parameters  $(\alpha, \lambda)$  and cdf  $F(z) = 1 - \left(\frac{\lambda}{\lambda+z}\right)^\alpha, z > 0$ .
- (iv) *Model 4*: Weibull distribution with parameters  $(c, \gamma)$  and cdf  $F(z) = 1 - e^{-cz^\gamma}, z > 0$ .
- (v) *Model 5*: Empirical distribution.

Recall that  $\mathbf{p}_5 = \frac{1}{n}\mathbf{1}$ . For all other models,  $\mathbf{p}_k$ 's are obtained by discretising the Maximum Likelihood fitted model. For example,

$$p_{ik} = F_k\left(\frac{x_{i+1} + x_i}{2}; \hat{\nu}\right) - F_k\left(\frac{x_i + x_{i-1}}{2}; \hat{\nu}\right),$$

for all  $i = 1, \dots, n, k \in \{1, 2, 3, 4\}$ ,

where by convention  $x_0 = -\infty$  and  $x_{n+1} = \infty$ . Moreover,  $\hat{\nu}$  is the Maximum Likelihood estimate of the unknown parameters.

The next step is to understand whether or not robust optimisation reduces the variability of the optimal decision. Thus,  $(\mathbf{y}_{wc}^*, P_{wc}^*)$ ,  $(\mathbf{y}_{wr}^*, P_{wr}^*)$ ,  $(\mathbf{y}_{AIC}^*, P_{AIC}^*)$  and  $(\mathbf{y}_{AICc}^*, P_{AICc}^*)$  are compared under three collections of candidate models denoted as  $\mathcal{M}_j, j \in \{2, 4, 5\}$ . In particular,  $\mathcal{M}_2 := \{1, 5\}$ ,  $\mathcal{M}_4 := \{1, 3, 4, 5\}$  and  $\mathcal{M}_5 := \{1, 2, 3, 4, 5\}$ . Recall that our sample is drawn from a LogNormal distribution model and for this reason the “true model”, i.e. Model 2, is purposely ruled out from  $\mathcal{M}_2$  and  $\mathcal{M}_4$ . Therefore, it would be interesting to understand the effect of reducing the model risk and analyse the robust optimal solutions under  $\mathcal{M}_2^* := \{2, 5\}$  and  $\mathcal{M}_4^* := \{2, 3, 4, 5\}$ . That is, Model 1 (that exhibits the lightest tail amongst all considered parametric models) is replaced by the “true model” (with a moderated light tail). For each of the model collection  $\mathcal{M}_j$  and  $\mathcal{M}_j^*, j \in \{2, 4, 5\}$  and  $l \in \{2, 4\}$ ,  $(\mathbf{y}_{wc}^*, P_{wc}^*)$  and  $(\mathbf{y}_{wr}^*, P_{wr}^*)$  are obtained by empirically solving the robust models (2.4) and (2.5) with  $\mathcal{M} = \mathcal{M}_j$  and  $\mathcal{M} = \mathcal{M}_j^*$ .

Before we explain the procedure of finding  $(\mathbf{y}_{AIC}^*, P_{AIC}^*)$  and  $(\mathbf{y}_{AICc}^*, P_{AICc}^*)$ , we briefly explain the AIC and AICc model selection. Given that a sample  $\mathbf{x}$  and a set of candidate probability models, the AIC value of Model  $k$  is calculated as  $AIC_k = 2q_k - 2Ln(\hat{L}_k)$ ,

where  $q_k$  is the number of parameters estimated and  $\hat{L}_k$  is the maximum value of the likelihood function of Model  $k$ . Under the AIC model selection criterion, the preferred Model  $k^*$  is the one that gives the smallest AIC value, i.e.  $k^* = \operatorname{argmin}_{k \in \mathcal{M}} AIC_k$ . On the other hand, the AICc value of Model  $k$  penalises the utility of each model for its complexity when the sample size  $n$  is not large, i.e.  $AICc_k = 2q_k \times \frac{n}{n-1-q_k} - 2Ln(\hat{L}_k)$ . Similarly, the preferred Model  $k^{**}$  under the AICc criterion is chosen to have the smallest AICc value, i.e.  $k^{**} = \operatorname{argmin}_{k \in \mathcal{M}} AICc_k$ . Finally,  $(\mathbf{y}_{AIC}^*, P_{AIC}^*)$  and  $(\mathbf{y}_{AICc}^*, P_{AICc}^*)$  are obtained by solving (2.4) with  $\mathcal{M} = \{k^*\}$  and  $\mathcal{M} = \{k^{**}\}$ , respectively.

Denote the underlying distribution of  $X$  as Model 0 that is equipped with its discretised probability vector  $\mathbf{p}_0$  obtainable as before:

$$p_{i0} = F_0\left(\frac{x_{i+1} + x_i}{2}\right) - F_0\left(\frac{x_i + x_{i-1}}{2}\right), \text{ for all } i = 1, \dots, n,$$

where  $F_0$  is the cdf of  $X$ , i.e. a LogNormal distribution with parameters defined earlier. Let  $(\mathbf{y}_T^*, P_T^*)$  be the optimal solution obtained by solving a non-robust version of Problem 2.4 with  $\mathcal{M} = \{0\}$  as given by  $\mathbf{p}_0^*$ . This optimal solution mimics the ideal optimal decision, since the “true” distribution is assumed to be known and thus, all possible robust methods are compared with the decision under Model 0. Clearly, the model risk induces uncertainty with the model choice and this issue is numerically experimented in the remaining part of this section. In order to compare various decisions, we need to measure the distance between the robust methods and the one obtained via Model 0. That is, each optimal contract  $\mathbf{y}_\xi^*$ , where  $\xi \in \{wc, wr, AIC, AICc\}$ , is compared to the benchmark optimal contract  $\mathbf{y}_T^*$  as follows:

$$\Delta_\xi = \sum_{i=1}^n |y_{i\xi}^* - y_{iT}^*| \times p_{i0} \text{ for all } \xi \in \{wc, wr, AIC, AICc\}.$$

Clearly, the smaller the value for  $\Delta_\xi$  is, the more robust of a decision is achieved. This criterion, further called *simple criterion*, assesses the choice of the optimal contract and it would be interesting to understand the possible drawbacks of those robust contracts, which may require increased premiums. A *composite criterion* would be needed when comparing  $(\mathbf{y}_w^*, P_w^*)$  to  $(\mathbf{y}_c^*, P_c^*)$  and is given by:

- a)  $(\mathbf{y}_w^*, P_w^*)$  is preferred and called “good scenario” if  $\Delta_w < \Delta_c$  and  $P_w^* - P_c^* < 10^{-2}$ ;
- b)  $(\mathbf{y}_c^*, P_c^*)$  is preferred and called “bad scenario” if  $\Delta_w > \Delta_c$  and  $P_w^* - P_c^* > -10^{-2}$ .

for any given  $w \in \{wc, wr\}$  and  $c \in \{AIC, AICc\}$ . Our numerical illustrations generate samples of size  $n \in \{25, 50, 100, 250\}$  for  $N = 500$  times and compare the robust optimal decisions to the AIC non-robust optimal decisions under the two criteria (simple and composite). Extensive numerical experiments (for various parametric models and sample sizes) have shown that the AIC and AICc-based optimal decisions lead to similar results and for this reason, only AIC results are further reported. That is, we display the number of “good” and “bad” scenarios, namely  $G_{w,AIC}$  and  $B_{w,AIC}$ , respectively, where  $w \in \{wc, wr\}$ .

We first examine the VaR<sub>0.75</sub>-based optimal solutions for the simple test, where only the robustness of the risk transfer is analysed. Table 6.1 shows those results when the “true model” is removed from two of the three candidate models, while Table 6.2 displays similar results when the LogNormal is always present amongst all potential parametric distributions.

Clearly, introducing the “true model” amongst the potential models, our robust methods are more efficient, but not sufficiently enough for various sample sizes; there is a marginal incentive to use our methods for small and medium sized samples. The results

**Table 6.1**  
Number of good and bad scenarios for VaR<sub>0.75</sub>-based scenarios within 500 samples of various sample sizes  $n$  and collection of candidate models  $\{\mathcal{M}_2, \mathcal{M}_4, \mathcal{M}_5\}$  under the simple criterion.

|              | $n = 25$        |                 |                 | $n = 50$        |                 |                 | $n = 100$       |                 |                 | $n = 250$       |                 |                 |
|--------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|              | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ |
| $G_{wc,AIC}$ | 222             | 217             | 225             | 196             | 195             | 195             | 143             | 143             | 143             | 67              | 67              | 68              |
| $B_{wc,AIC}$ | 278             | 283             | 275             | 304             | 305             | 305             | 357             | 357             | 357             | 433             | 433             | 432             |
| $G_{wr,AIC}$ | 258             | 257             | 250             | 255             | 246             | 236             | 200             | 185             | 182             | 125             | 110             | 109             |
| $B_{wr,AIC}$ | 242             | 243             | 250             | 245             | 254             | 264             | 300             | 315             | 318             | 375             | 390             | 391             |

**Table 6.2**  
Number of good and bad scenarios for VaR<sub>0.75</sub>-based scenarios within 500 samples of various sample sizes  $n$  and collection of candidate models  $\{\mathcal{M}_2^*, \mathcal{M}_4^*, \mathcal{M}_5\}$  under the simple criterion.

|              | $n = 25$        |                   |                   | $n = 50$        |                   |                   | $n = 100$       |                   |                   | $n = 250$       |                   |                   |
|--------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|
|              | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ |
| $G_{wc,AIC}$ | 222             | 225               | 279               | 196             | 179               | 291               | 143             | 134               | 310               | 67              | 47                | 278               |
| $B_{wc,AIC}$ | 278             | 275               | 221               | 304             | 321               | 209               | 357             | 366               | 189               | 433             | 453               | 222               |
| $G_{wr,AIC}$ | 258             | 260               | 304               | 255             | 247               | 318               | 200             | 210               | 345               | 125             | 95                | 345               |
| $B_{wr,AIC}$ | 242             | 240               | 196               | 245             | 253               | 182               | 300             | 290               | 155               | 375             | 405               | 155               |

**Table 6.3**  
Number of good and bad scenarios for non-comonotonic CVaR<sub>0.75</sub>-based scenarios within 500 samples of various sample sizes  $n$  and collection of candidate models  $\{\mathcal{M}_2, \mathcal{M}_4, \mathcal{M}_5\}$  under the simple criterion.

|                | $n = 25$        |                 |                 | $n = 50$        |                 |                 | $n = 100$       |                 |                 | $n = 250$       |                 |                 |
|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ |
| $G_{wc,AIC}$   | 337             | 349             | 353             | 307             | 331             | 338             | 328             | 331             | 329             | 307             | 308             | 291             |
| $B_{wc,AIC}$   | 163             | 151             | 147             | 193             | 169             | 162             | 172             | 169             | 171             | 193             | 192             | 209             |
| $G_{wr,AIC}$   | 336             | 339             | 339             | 292             | 304             | 305             | 287             | 302             | 268             | 222             | 240             | 204             |
| $B_{wr,AIC}$   | 164             | 161             | 161             | 208             | 196             | 195             | 213             | 198             | 232             | 278             | 260             | 296             |
| $G_{wcvr,AIC}$ | 322             | 323             | 323             | 305             | 327             | 336             | 305             | 326             | 324             | 305             | 304             | 313             |
| $B_{wcvr,AIC}$ | 178             | 177             | 177             | 195             | 172             | 164             | 195             | 174             | 176             | 195             | 196             | 187             |

**Table 6.4**  
Number of good and bad scenarios for non-comonotonic CVaR<sub>0.75</sub>-based scenarios within 500 samples of various sample sizes  $n$  and collection of candidate models  $\{\mathcal{M}_2^*, \mathcal{M}_4^*, \mathcal{M}_5\}$  under the simple criterion.

|                | $n = 25$        |                   |                   | $n = 50$        |                   |                   | $n = 100$       |                   |                   | $n = 250$       |                   |                   |
|----------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|
|                | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ |
| $G_{wc,AIC}$   | 337             | 340               | 354               | 307             | 308               | 306               | 328             | 328               | 315               | 307             | 309               | 305               |
| $B_{wc,AIC}$   | 163             | 160               | 146               | 193             | 192               | 194               | 172             | 172               | 185               | 193             | 191               | 195               |
| $G_{wr,AIC}$   | 336             | 340               | 330               | 292             | 299               | 270               | 287             | 295               | 292               | 222             | 234               | 265               |
| $B_{wr,AIC}$   | 164             | 160               | 170               | 208             | 201               | 230               | 213             | 205               | 208               | 278             | 266               | 235               |
| $G_{wcvr,AIC}$ | 322             | 312               | 306               | 305             | 301               | 289               | 305             | 324               | 302               | 305             | 300               | 321               |
| $B_{wcvr,AIC}$ | 178             | 187               | 194               | 195             | 199               | 211               | 195             | 176               | 198               | 195             | 200               | 179               |

**Table 6.5**  
Number of good and bad scenarios for non-comonotonic CVaR<sub>0.75</sub>-based scenarios within 500 samples of various sample sizes  $n$  and collection of candidate models  $\{\mathcal{M}_2, \mathcal{M}_4, \mathcal{M}_5\}$  under the composite criterion.

|                | $n = 25$        |                 |                 | $n = 50$        |                 |                 | $n = 100$       |                 |                 | $n = 250$       |                 |                 |
|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ |
| $G_{wc,AIC}$   | 333             | 345             | 351             | 307             | 331             | 338             | 328             | 331             | 329             | 307             | 308             | 291             |
| $B_{wc,AIC}$   | 158             | 146             | 140             | 193             | 169             | 162             | 172             | 169             | 171             | 193             | 192             | 209             |
| $G_{wr,AIC}$   | 332             | 335             | 336             | 292             | 304             | 305             | 287             | 302             | 268             | 222             | 240             | 204             |
| $B_{wr,AIC}$   | 157             | 154             | 152             | 208             | 196             | 194             | 213             | 198             | 231             | 278             | 259             | 296             |
| $G_{wcvr,AIC}$ | 316             | 317             | 316             | 305             | 327             | 336             | 305             | 326             | 324             | 305             | 304             | 313             |
| $B_{wcvr,AIC}$ | 176             | 174             | 177             | 195             | 172             | 164             | 195             | 174             | 176             | 195             | 196             | 187             |

under the composite criterion (not reported) lead to a similar conclusion. This is not surprising, since the VaR risk measure is quite robust (see Cont et al., 2010) in the sense that the whole sample could be contaminated, but not a single value, and still have the same estimate. This peculiar behaviour of this tail risk measure explains why our methods are not recommended for VaR-based decisions.

Next, we turn our attention to another tail risk measure, namely CVaR<sub>0.75</sub>. Our results for various sample sizes  $n$  and model collections are presented in Tables 6.3 and 6.4 for the simple criterion, while Tables 6.5 and 6.6 replicate similar results for the

composite criterion. The first two rows of Tables 6.3–6.6 are computed via the LP formulation from Theorem 4.1, while the results of the third and fourth rows are based on the LP reformulation from Proposition 4.1. The last two rows are obtained by optimising the WCvAr risk measure, as defined in (2.8).

There is an overwhelming empirical evidence that our worst-case scenario method performs uniformly better than the WCvAr robust method from Zhu and Fukushima (2009) under both criteria, simple and composite, for any sample size. This could be explained by the fact WCvAr is a more conservative risk measure than our proposed robust risk measures. It is interesting to note

**Table 6.6**

Number of good and bad scenarios for non-comonotonic CVaR<sub>0.75</sub>-based scenarios within 500 samples of various sample sizes  $n$  and collection of candidate models  $\{\mathcal{M}_2^*, \mathcal{M}_4^*, \mathcal{M}_5\}$  under the composite criterion.

|                | $n = 25$        |                   |                   | $n = 50$        |                   |                   | $n = 100$       |                   |                   | $n = 250$       |                   |                   |
|----------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|
|                | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ |
| $G_{wc,AIC}$   | 333             | 336               | 354               | 307             | 308               | 306               | 328             | 328               | 315               | 307             | 309               | 305               |
| $B_{wc,AIC}$   | 158             | 155               | 137               | 193             | 192               | 194               | 172             | 172               | 185               | 193             | 191               | 195               |
| $G_{wr,AIC}$   | 332             | 336               | 330               | 292             | 299               | 270               | 287             | 295               | 292               | 222             | 234               | 265               |
| $B_{wr,AIC}$   | 157             | 152               | 155               | 208             | 201               | 227               | 213             | 205               | 208               | 278             | 266               | 235               |
| $G_{wcvr,AIC}$ | 316             | 306               | 299               | 305             | 301               | 289               | 305             | 324               | 302               | 305             | 300               | 321               |
| $B_{wcvr,AIC}$ | 176             | 185               | 192               | 195             | 199               | 211               | 195             | 176               | 198               | 195             | 200               | 179               |

**Table 6.7**

Number of good and bad scenarios for comonotonic CVaR<sub>0.75</sub>-based scenarios within 500 samples of various sample sizes  $n$  and collection of candidate models  $\{\mathcal{M}_2, \mathcal{M}_4, \mathcal{M}_5\}$  under the simple criterion.

|                | $n = 25$        |                 |                 | $n = 50$        |                 |                 | $n = 100$       |                 |                 | $n = 250$       |                 |                 |
|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ |
| $G_{wc,AIC}$   | 259             | 247             | 264             | 277             | 282             | 270             | 315             | 333             | 307             | 301             | 313             | 284             |
| $B_{wc,AIC}$   | 241             | 253             | 236             | 223             | 217             | 230             | 185             | 167             | 193             | 199             | 187             | 216             |
| $G_{wr,AIC}$   | 307             | 307             | 318             | 304             | 305             | 302             | 283             | 278             | 263             | 221             | 223             | 206             |
| $B_{wr,AIC}$   | 193             | 193             | 182             | 196             | 195             | 198             | 217             | 222             | 237             | 279             | 277             | 294             |
| $G_{wcvr,AIC}$ | 259             | 271             | 234             | 264             | 277             | 256             | 277             | 296             | 265             | 282             | 297             | 267             |
| $B_{wcvr,AIC}$ | 241             | 229             | 265             | 236             | 223             | 244             | 223             | 204             | 235             | 218             | 203             | 233             |

**Table 6.8**

Number of good and bad scenarios for comonotonic CVaR<sub>0.75</sub>-based scenarios within 500 samples of various sample sizes  $n$  and collection of candidate models  $\{\mathcal{M}_2^*, \mathcal{M}_4^*, \mathcal{M}_5\}$  under the simple criterion.

|                | $n = 25$        |                   |                   | $n = 50$        |                   |                   | $n = 100$       |                   |                   | $n = 250$       |                   |                   |
|----------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|
|                | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ |
| $G_{wc,AIC}$   | 259             | 250               | 230               | 277             | 274               | 252               | 315             | 311               | 290               | 301             | 306               | 298               |
| $B_{wc,AIC}$   | 241             | 250               | 270               | 223             | 226               | 248               | 185             | 189               | 210               | 199             | 194               | 202               |
| $G_{wr,AIC}$   | 307             | 311               | 308               | 304             | 313               | 305               | 283             | 286               | 319               | 221             | 231               | 299               |
| $B_{wr,AIC}$   | 193             | 189               | 192               | 196             | 187               | 195               | 217             | 214               | 181               | 279             | 269               | 201               |
| $G_{wcvr,AIC}$ | 259             | 249               | 215               | 264             | 250               | 202               | 277             | 273               | 239               | 282             | 285               | 233               |
| $B_{wcvr,AIC}$ | 241             | 251               | 285               | 236             | 250               | 298               | 223             | 227               | 261               | 218             | 215               | 267               |

**Table 6.9**

Number of good and bad scenarios for comonotonic CVaR<sub>0.75</sub>-based scenarios within 500 samples of various sample sizes  $n$  and collection of candidate models  $\{\mathcal{M}_2, \mathcal{M}_4, \mathcal{M}_5\}$  under the composite criterion.

|                | $n = 25$        |                 |                 | $n = 50$        |                 |                 | $n = 100$       |                 |                 | $n = 250$       |                 |                 |
|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ | $\mathcal{M}_5$ | $\mathcal{M}_4$ | $\mathcal{M}_2$ |
| $G_{wc,AIC}$   | 253             | 240             | 258             | 276             | 281             | 269             | 315             | 333             | 307             | 301             | 313             | 284             |
| $B_{wc,AIC}$   | 239             | 252             | 235             | 223             | 217             | 230             | 185             | 167             | 193             | 199             | 187             | 216             |
| $G_{wr,AIC}$   | 299             | 299             | 312             | 302             | 303             | 301             | 283             | 278             | 263             | 221             | 223             | 206             |
| $B_{wr,AIC}$   | 191             | 191             | 180             | 196             | 195             | 198             | 217             | 222             | 237             | 279             | 277             | 294             |
| $G_{wcvr,AIC}$ | 255             | 267             | 233             | 263             | 276             | 255             | 277             | 296             | 265             | 282             | 297             | 267             |
| $B_{wcvr,AIC}$ | 237             | 225             | 259             | 236             | 223             | 244             | 223             | 204             | 235             | 218             | 203             | 233             |

**Table 6.10**

Number of good and bad scenarios for comonotonic CVaR<sub>0.75</sub>-based scenarios within 500 samples of various sample sizes  $n$  and collection of candidate models  $\{\mathcal{M}_2^*, \mathcal{M}_4^*, \mathcal{M}_5\}$  under the composite criterion.

|                | $n = 25$        |                   |                   | $n = 50$        |                   |                   | $n = 100$       |                   |                   | $n = 250$       |                   |                   |
|----------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|-----------------|-------------------|-------------------|
|                | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ | $\mathcal{M}_5$ | $\mathcal{M}_4^*$ | $\mathcal{M}_2^*$ |
| $G_{wc,AIC}$   | 253             | 243               | 227               | 276             | 273               | 252               | 315             | 311               | 290               | 301             | 306               | 298               |
| $B_{wc,AIC}$   | 239             | 249               | 266               | 223             | 226               | 248               | 185             | 189               | 210               | 199             | 194               | 202               |
| $G_{wr,AIC}$   | 299             | 304               | 304               | 302             | 311               | 305               | 283             | 286               | 319               | 221             | 231               | 299               |
| $B_{wr,AIC}$   | 191             | 188               | 189               | 196             | 187               | 195               | 217             | 214               | 181               | 279             | 269               | 201               |
| $G_{wcvr,AIC}$ | 255             | 245               | 215               | 263             | 249               | 202               | 277             | 273               | 239               | 282             | 285               | 233               |
| $B_{wcvr,AIC}$ | 237             | 247               | 278               | 236             | 250               | 298               | 223             | 227               | 261               | 218             | 215               | 267               |

that the worst-case regret method tends to under-perform both worst-case methods, but we believe that this due to the fact CVaR is a tail risk measure.

Tables 6.7–6.10 are the replica of Tables 6.3–6.5, where the set of feasible solutions is reduced such that the insurance contracts are assumed to be comonotone. That is, the first two rows of

Tables 6.7–6.10 are computed as explained in Remark 3.2, while the results of the third and fourth rows are based on an LP formulation similar to the one from Proposition 3.1. As before, the last two rows are obtained by optimising the WCVaR risk measure, as defined in (2.8), but adding the comonotonicity constraint. Restricting our optimisation to comonotone contracts does not

change our results, but we observe a loss of power amongst all three robust methods, which could be explained by the fact that an additional constraint increases the complexity of the problem. The general conclusions do not change and there is clear evidence to recommend our worst-case scenario method that outperforms the WCVaR robust method from Zhu and Fukushima (2009) and our worst-case regret method under both criteria.

As a final remark, it is worth mentioning that applying Theorem 5.1 to our robust methods, all numerical results remain unchanged. Therefore, the power of results are similar to those displayed in the section, which suggests that one should use our worst-case method in conjunction with Theorem 5.1 in order to obtain a robust insurance contract that economically is viable to both insurance players.

## 7. Conclusions

The VaR and CVaR-based optimal insurance contract has been investigated under uncertainty, where the model risk is taken into account. This source of uncertainty is considered by incorporating multiple plausible models that the decision-maker would have available via estimation, proxy models or expert opinion consultation. Model risk always represents an important source of uncertainty in risk modelling and it is more pronounced when data scarcity is present. Our aim has been to provide a robust decision and not to produce a distribution robust method of the underlying insurance risk. Two robust methods are proposed, namely the Worst-case and Worst-regret. Our numerical results have shown that our Worst-case method outperforms the Worst-regret method for CVaR-based decisions. Moreover, our Worst-case method proved to be more robust than the Worst-case CVaR method proposed by Zhu and Fukushima (2009). Unfortunately, the VaR-based decisions are not efficiently robustified for all sample sizes by neither methods proposed in this paper, though encouraging results are obtained for small samples. Another achievement of this paper is related to the well-known caveat in robust optimisation that optimal decision may be economically unacceptable. That is, the optimal contract may be not efficient in the Pareto optimality sense. We resolve this issue by providing a simple numerical method that allows one to identify an optimal Pareto and robust decision that is (numerically) shown to be efficient for reducing the model risk.

## Acknowledgments

K. C. Cheung, acknowledges the financial support from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 17324516).

## References

- Acerbi, C., & Székely, B. (2014). Backtesting expected shortfall. *Risk magazine*.
- Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7), 1487–1503.
- Appa, G. (2002). On the uniqueness of solutions to linear programs. *Journal of the Operational Research Society*, 53(10), 1127–1132.
- Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53(3), 941–973.
- Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203–228.
- Asimit, A. V., Badescu, A. M., & Tsanakas, A. (2013a). Optimal risk transfers in insurance groups. *European Actuarial Journal*, 3, 159–190.
- Asimit, A. V., Badescu, A. M., & Verdonck, T. (2013b). Optimal risk transfer under quantile-based risk measures. *Insurance: Mathematics and Economics*, 53, 252–265.
- Asimit, V., Gao, T., Hu, J., & Kim, E. S. (2017). Optimal risk transfer: A numerical optimisation approach. Available at: [http://papers.ssrn.com/abstract\\_id=2797562](http://papers.ssrn.com/abstract_id=2797562).
- Balbás, A., Balbás, B., Balbás, R., & Heras, A. (2015). Optimal reinsurance under risk and uncertainty. *Insurance: Mathematics and Economics*, 60, 61–74.
- Balbás, A., Balbás, B., & Heras, A. (2009). Optimal reinsurance with general risk measures. *Insurance: Mathematics and Economics*, 44, 374–384.
- Balbás, A., Balbás, B., & Heras, A. (2011). Stable solutions for optimal reinsurance problems involving risk measures. *European Journal of Operational Research*, 214, 796–804.
- Bellini, F., & Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance*, 15(5), 725–733.
- Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization*. New Jersey: Princeton University Press.
- Ben-Tal, A., & Nemirovski, A. (2008). Selected topics in robust convex optimization. *Mathematical Programming*, 112(1), 125–158.
- Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, 53(3), 464–501.
- Borch, K. (1960). An attempt to determine the optimum amount of stop loss reinsurance. In *Transactions of the 16th international congress of actuaries I* (pp. 597–610).
- Cai, J., & Tan, K. S. (2007). Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measure. *ASTIN Bulletin*, 37(1), 93–112.
- Cai, J., & Wei, W. (2012). Optimal reinsurance with positively dependent risks. *Insurance: Mathematics and Economics*, 50, 57–63.
- Cai, J., & Weng, C. G. (2016). Optimal reinsurance with expectile. *Scandinavian Actuarial Journal*, 7, 624–645.
- Calder, A., Couper, A., & Lo, J. (2012). Catastrophe model blending techniques and governance. The actuarial profession. URL <https://www.actuaries.org.uk/documents/brian-hey-prize-catastrophe-model-blending>.
- Cheung, K. C., & Lo, A. (2017). Characterizations of optimal reinsurance treaties: a cost-benefit approach. *Scandinavian Actuarial Journal*, 2017(1), 1–28.
- Cheung, K. C., Sung, K. C. J., Yam, S. C. P., & Yung, S. P. (2014). Optimal reinsurance under general law-invariant risk measures. *Scandinavian Actuarial Journal*, 1, 72–91.
- Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3), 519–547.
- Cont, R., Deguest, R., & Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10(6), 593–606.
- Delbaen, F., Bellini, F., Bignozzi, V., & Ziegel, J. F. (2016). On risk measures with the CxLS property. *Finance and Stochastics*, 20(2), 433–453.
- Dhaene, J., Kukush, A., Linders, D., & Tang, Q. (2012). Remarks on quantiles and distortion risk measures. *European Actuarial Journal*, 2(2), 319–328.
- El Ghaoui, L., Oks, M., & Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4), 543–556.
- Emmer, S., Kratz, M., & 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., & Ziegel, J. F. (2016). Higher order elicibility and Osband's principle. *The Annals of Statistics*, 44(4), 1680–1707.
- Gabrel, V., Murat, C., & Thiele, A. (2014). Recent advances in robust optimization: An overview. *European Journal of Operational Research*, 235(3), 471–483.
- Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2), 141–153.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494), 746–762.
- Güler, O. (2010). *Foundations of optimization*. New York: Springer.
- Hampel, F. R. (1968). *Contribution to the theory of robust estimation*, Ph.D Thesis. University of California, Berkeley.
- Huang, D., Zhu, S., Fabozzi, F. J., & Fukushima, M. (2010). Portfolio selection under distributional uncertainty: A relative robust CVaR approach. *European Journal of Operational Research*, 203(1), 185–194.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1), 73–101.
- Huberman, G., Mayers, D., & Smith, C. W., Jr. (1983). Optimal insurance policy indemnity schedules. *The Bell Journal of Economics*, 14(2), 415–426.
- Iancu, D. A., & Trichakis, N. (2014). Pareto efficiency in robust optimisation. *Management Science*, 60(1), 130–147.
- Kakouris, I., & Rustem, B. (2014). Robust portfolio optimization with copulas. *European Journal of Operational Research*, 235(1), 28–37.
- Kaluszka, M. (2005). Optimal reinsurance under convex principles of premium calculation. *Insurance: Mathematics and Economics*, 36(3), 375–398.
- Kaluszka, M., & Okolewski, A. (2008). An extension of arrows result on optimal reinsurance contract. *Journal of Risk and Insurance*, 75(2), 275–288.
- Kiesel, R., Rühlicke, R., Stahl, G., & Zheng, J. (2016). The Wasserstein metric and robustness in risk management. *Risks*, 4(3), 32.
- Koch-Medina, P., & Munari, C. (2016). Unexpected shortfalls of expected shortfall: Extreme default profiles and regulatory arbitrage. *Journal of Banking & Finance*, 62, 141–151.
- Krätschmer, V., Schied, A., & Zähle, H. (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastics*, 18(2), 271–295.
- Landsberger, M., & Meiljson, I. (1994). Co-monotone allocations, Bickel-Lehmann dispersion and the Arrow-Pratt measure of risk aversion. *Annals of Operations Research*, 52(2), 97–106.
- Ludkovski, M., & Rüschendorf, L. (2008). On comonotonicity of Pareto optimal allocations. *Statistics and Probability Letters*, 78(10), 1181–1188.
- Mangasarian, O. L. (1979). Uniqueness of solution in linear programming. *Linear Algebra and its Applications*, 25, 151–162.
- Polak, G. G., Rogers, D. F., & Sweeney, D. J. (2010). Risk management strategies via minimax portfolio optimization. *European Journal of Operational Research*, 207(1), 409–419.
- Rockafeller, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2, 21–41.

- Song, Y., & Yan, J. A. (2009). Risk measures with comonotonic subadditivity or convexity and respecting stochastic orders. *Insurance: Mathematics and Economics*, 45(3), 459–465.
- SPAN (1995). *Standard portfolio analysis of risk*. Chicago: Chicago Mercantile Exchange.
- Wang, S., Young, V. R., & Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21(2), 173–183.
- Young, V. R. (2004). Premium principles. *Encyclopedia of actuarial science*. New York: Wiley.
- Zhu, S., & 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., & Rustem, B. (2013). Worst-case value-at-risk of nonlinear portfolios. *Management Science*, 59(1), 172–188.