OPTIMAL ROBUST INSURANCE WITH A FINITE UNCERTAINTY SET

Alexandru V. Asimit

Cass Business School, City, University of London, London EC1Y 8TZ, United Kingdom. E-mail: asimit@city.ac.uk

Junlei Hu

Department of Mathematical Sciences, University of Essex, Colchester CO4 3SQ United Kingdom. E-mail: j.hu@essex.ac.uk

Yuantao Xie[1]

School of Insurance and Economics, University of International Business and Economics, Beijing, 100029, China. E-mail: xieyuantao@uibe.edu.cn

March 29, 2019

Abstract

Decision-makers who usually face model/parameter risk may prefer to act prudently by identifying optimal contracts that are robust to such sources of uncertainty. In this paper, we tackle this issue under a finite *uncertainty set* that contains a number of probability models that are candidates for the "true", but unknown model. Various robust optimisation models are proposed, some of which are already known in the literature, and we show that all of them can be efficiently solved via Second Order Conic Programming (SOCP). Numerical experiments are run for various risk preference choices and it is found that for relatively large sample size, the modeler should focus on finding the best possible fit for the unknown probability model in order to achieve the most robust decision. If only small samples are available, then the modeler should consider two robust optimisation models, namely the Weighted Average Model or Weighted Worst-case Model, rather than focusing on statistical tools aiming to estimate the probability model. Amongst those two, the better choice of the robust optimisation model depends on how much interest the modeler puts on the tail risk when defining its objective function. These findings suggest that one should be very careful when robust optimal decisions are sought in the sense that the modeler should first understand the features of its objective function and the size of

[1]Corresponding author. Phone: +861064494702. Fax: +861064495024.

the available data, and then to decide whether robust optimisation or statistical inferences is the best practical approach.

Keywords and phrases: Optimal reinsurance, Risk measure, Robust optimisation, Second order conic programming, Uncertainty modelling.

1. Introduction

The seminal works by Borch (1960) and Arrow (1963) mark the beginning of the theory of optimal insurance/reinsurance in the field of actuarial science. The same problem is known as the insurance demand problem in the field of insurance economics. In the last 50 years, many research outputs have contributed into these fields of research by identifying the optimal insurance/reinsurance contracts under various risk preferences. Examples outside the Expected Utility Theory are numerous; for example, risk measure-based models have been studied by Cai *et al.* (2008), Balbás *et al.* (2009 and 2011), Chi and Tan (2011), Asimit *et al.* (2013 and 2015), Cheung *et al.* (2014), Lu *et al.* (2014) and Cai and Weng (2016), where *Value-at-Risk (VaR)* and *Conditional-Value-at-Risk (CVaR)* based decisions are the focal interest, since these particular risk preferences are easy to interpret and are the most common in the insurance sector.

The majority of the contributions from the existing literature assumes that the model specifications are completely known, which purposely removes the model and parameter risks – the risk of choosing a "wrong" model or the risk of choosing the "right" parametric model with the "wrong" parameter values/estimates. Such risks are not of great concern when modelling is based on high frequency data or simply, when large samples are available. Unfortunately, data scarcity is a common feature of insurance data, which increases the uncertainty within the modelling process and making any risk measurement to be highly sensitive. Therefore, the standard statistical methods that aim to identify the "best" model fail to provide a reasonable answer. Solutions to incorporate the model/parameter risks are available in the statistical literature, for example parametric and non-parametric bootstrapping. Moreover, there exists a large number of literatures in the field of actuarial science on finding robust worst-case risk measures, which is reviewed and extended in Goovaerts *et al.* (2011). Any of these are possible whenever a simple risk measurement is performed. This is no longer the case when the main aim is to find the best strategy within an optimisation problem, where finding the "best" model does not guarantee a robust decision, which is the main aim of the modelling process. A standard way to achieve this is to use the method of robust optimisation; comprehensive surveys could be found, for example, in Ben-Tal and Nemirovski (2002 and 2008), Ben-Tal *et al.* (2009), Bertsimas *et al.* (2011) and

Gabrel *et al.* (2014), while applications to the optimal insurance literature could be found in Balbás *et al.* (2015) and Asimit *et al.* (2017).

In this paper, we aim to identify the optimal insurance contract using a robust optimisation model with a finite uncertainty set. That is, the modeler does not know which probability model is appropriate and the optimal decision is produced by incorporating the risk measurements under all (but in a finite number) of the possible probability models. That is, the uncertainty set is constructed over a finite number of models as in Zhu and Fukushima (2009), Huang *et al.* (2010) and Asimit *et al.* (2017), where the first two papers considered a convex hull of the candidate models. This approach leads to a large uncertainty set that may be detrimental to the robust optimal decision and therefore, it would be better to consider a non-convex uncertainty set that is purely composed of the possible models as explained in Asimit *et al.* (2017). We extend this idea by investigating various robust optimisation formulations and try to understand the effect over the robustness of the optimal decision, which is in fact the main aim of robust optimisation. In order to be more explicit, all formulations are explored within the context of optimal insurance, but any application would lead to similar investigations. Our model assumes homogeneous multiple beliefs with respect to the distribution of the buyer's initial exposure. Distributional uncertainty could be perceived in different ways by the insurance buyer and insurance seller, but this would not change our mathematical formulations if the two parties have different beliefs about the distributional ambiguity set of X. Adverse selection is the classical example of asymmetric information between the buyer and seller that would justify the distributional ambiguity. The risk modelling power and experience are also related to the market size of an insurance company, which explains why some insurance players have competition advantage when consuming the data from different other sources. For example, machine learning techniques are good candidates to extract valuable information form data that are not obviously informative to explain the risk in question via the classical actuarial techniques; this includes combining various databases via dimension reduction methods and cluster analysis, which would help to enrich the risk experience of the insurer player that has enhanced analytic capabilities. A recent paper of Ghossoub (2019) also discusses the impact of heterogeneous beliefs over the Pareto set for a nice and tractable model. In contrast, our approach relies on numerical optimisation to characterise the Pareto set under any finite set of homogeneous multiple beliefs, but our aim is also to explain how these uncertainty sets could help to make the decision more robust in the presence of (distributional) model error given that the uncertainty sets could have an unknownable impact over actual decision.

The performance of our robust optimisation models are empirically evaluated via solving *Second Order Conic Programming (SOCP)* instances, which can be efficiently solved. SOCP problems are convex optimisation problems in which a linear objective function is minimised over the intersection of an affine set and the product of second-order (quadratic) cones. The well-known *Linear Programming (LP)*, *Quadratically Constrained Linear Programming (QCLP)* and *Quadratically Constrained Quadratic Programming (QCQP)* are SOCP examples (for details, see Alizadeh and Goldfarb, 2003). SOCP is a popular numerical method for engineering applications (for example, see Lobo *et al.*, 1998), robust portfolio optimisation (for example, see Satchell, 2010) or for actuarial/insurance applications (for example, see Tan and Weng, 2014, Asimit and Boonen, 2018 and Asimit *et al.*, 2018). The main reason behind the popularity of SOCP formulations is given by its computational efficiency. A number of efficient primal-dual interior-point methods for solving SOCP problems have been studied and developed in the literature. For example, Lobo *et al.* (1998) gives a worst-case theoretical analysis showing that the required number of iterations grow at most as the square root of the problem size. Therefore, by casting our robust optimisation models as SOCP problems, we are able to efficiently obtain the optimal solutions using SOCP solvers.

The paper is organised as follows: Section 2 explains the robust optimisation formulations, whose empirical formulations are discussed in Sections 3; extensive numerical examples are given in Section 4 that evaluates the quality of our robust solutions by comparing to some classic non-robust optimisation solutions; conclusions and all proofs are provided in Section 5 and 6, respectively.

2. Problem Formulation

2.1. **Standard Robust Optimisation Formulations.** Robust optimisation is widely recognised as an efficient method to incorporate the uncertainty with the model assumptions in an optimisation problem. If random variables are included in the objective function, then the parameter/model risks represent the uncertainty that one should take into account in order to create a robust optimal decision. Transforming information into knowledge, by means of finding an optimal decision that is less sensitive to the model inputs, is possible if the actual optimisation is performed over an uncertainty set. This set comprises of reasonable information available regarding the model parameters and/or competitive models that are considered realistic or common/good practice within the sector or profession. Specifically, the objective is to optimise $f(\cdot;\boldsymbol{\omega}) : \mathcal{A} \rightarrow \Re$ with $\mathcal{A}$ being a convex, where both are sensitive to the choice of model inputs.

The standard *worst-case (wc)* robust optimisation formulation is given by:

$$\min_{\mathbf{t}\in\mathcal{A}} \sup_{\boldsymbol{\omega}\in\mathcal{U}} f(\mathbf{t};\boldsymbol{\omega}), \tag{2.1}$$

where $\mathcal{U}$ is the uncertainty set that best describes the entire spectrum of model specifications. Given that the optimal decision is very sensitive to the model choice, any change in model inputs would possibly massively influence the optimum. Therefore, effectiveness may not necessarily be achieved by choosing the "best possible" model choice, which carries its own level of uncertainty, and robust optimisation is precisely created to help with producing robust decision. Note that our discrete and finite uncertainty set $\mathcal{U}$ is chosen to explain the model error faced by the decision-maker. Continuous uncertainty sets are also possible and are mathematically appealing and allows one to elaborate complex mathematical explorations. If the main objective is to deal with parameter uncertainty, then the uncertainty set is set around one reference probability model and the uncertainty could typically be described via *hyper-boxes*, *polytopes* or *ellipsoids*. Recall that the hyper-box uncertainty sets are sometimes known as the *interval uncertainty set*. More detailed analyses and evaluations on the performance of robust optimisation using finite and infinite uncertainty sets can be found in Ben-Tal and Nemirovski (2002 and 2008), Ben-Tal *et al.* (2009), Zymler *et al.* (2013) and Chassein and Goerigk (2016). The uncertainty sets from Zhu and Fukushima (2009) and Huang *et al.* (2010) are continuous and in fact, are the convex hull version of our discrete and finite uncertainty set choice. Note that this approach produces a large uncertainty set that may affect the robustness of the optimal decision, which is in fact the main purpose of the robust optimisation; for details, see Asimit *et al.* (2017) where it is shown that robust optimisation models with a finite uncertainty set tend to produce robust optimal solutions that are closer to the 'true' optimal solution. For these reasons, we have decided to consider this discrete and finite uncertainty set. Specifically, if $\mathcal{U} = \{\boldsymbol{\omega}_k, k \in \mathcal{M}\}$, where $\mathcal{M} := \{1, 2, \ldots, m\}$, then (2.1) becomes

$$\min_{\mathbf{t}\in\mathcal{A}} \max_{k\in\mathcal{M}} f(\mathbf{t};\boldsymbol{\omega}_k). \tag{2.2}$$

An alternative robust representation, namely the *worst-regret (wr)*-type, appears in the recent literature and its formulation is given by:

$$\min_{\mathbf{t}\in\mathcal{A}} \max_{k\in\mathcal{M}} f(\mathbf{t};\boldsymbol{\omega}_k) - f_k^*, \quad \textit{where} \quad f_k^* = \min_{\mathbf{t}\in\mathcal{A}} f(\mathbf{t};\boldsymbol{\omega}_k) \quad \text{for all } k \in \mathcal{M}. \tag{2.3}$$

For further details, see Huang *et al.* (2010) and Asimit *et al.* (2017). A Bayesian-type representation would be to average each possible model by allocating various weights to every single model according to the prior knowledge that the modeler might have. That is, with some given scalars λ_k, we have the following *weighted average (wa) type* robust problem

$$\min_{\mathbf{t}\in\mathcal{A}} \sum_{k\in\mathcal{M}} \lambda_k f(\mathbf{t};\boldsymbol{\omega}_k), \tag{2.4}$$

where $\boldsymbol{\lambda} \geq \mathbf{0}$ and $\mathbf{1}^T\boldsymbol{\lambda} = 1$. When the weights are all equal, the robust problem is labelled as *additive-type (ad).*

A robust risk measurement that has not been discussed in the literature is the following *weighted worst-case (wwc) scenario* type

$$f\big(\mathbf{t};\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_m,l\big) := \frac{1}{l}\sum_{i=1}^{l} f^{(i)}\big(\mathbf{t};\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_m,l\big),\ l\in\mathcal{M}, \tag{2.5}$$

where $f^{(i)}(\cdot;\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_m,l)$ is the i^{th} upper order statistics of $\big\{f(\cdot;\boldsymbol{\omega}_k), k\in\mathcal{M}\big\}$, i.e.

$$f^{(i)}\big(\mathbf{t};\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_m,l\big) = f\big(\mathbf{t};\boldsymbol{\omega}_{\sigma(i)}\big) \quad \text{such that} \quad f\big(\mathbf{t};\boldsymbol{\omega}_{\sigma(1)}\big) \geq \ldots \geq f\big(\mathbf{t};\boldsymbol{\omega}_{\sigma(l)}\big)$$

with σ being a permutation of $\mathcal{M}$. Essentially, the decision maker evaluates the model uncertainty as a weighted average of some higher tier risk levels that are measured over all possible assumptions assumed to be equally likely to occur. Note that any weighted worst-case scenario (2.5) is less conservative than (2.2) for any given l. Now, for any $\mathbf{t}\in\mathcal{A}$, (2.5) could be reformulated in the following fashion

$$f\big(\mathbf{t};\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_m,l\big) = \min_{s\in\Re}\left\{s + \frac{1}{l}\sum_{i\in\mathcal{M}}\Big(f\big(\mathbf{t};\boldsymbol{\omega}_i\big) - s\Big)_+\right\}, \quad \text{where} \quad (t)_+ = \max(t,0). \tag{2.6}$$

It is not difficult to obtain the result from (2.6) and therefore, let $\{a_k,\ k\in\mathcal{M}\}$ be a finite set. Therefore, one needs to show that

$$\frac{1}{l}\sum_{i=1}^{l} a^{(i)} = \min_{s\in\Re}\left\{s + \frac{1}{l}\sum_{i\in\mathcal{M}}\big(a_i - s\big)_+\right\},$$

where $a^{(i)}$ represents the i^{th} upper order statistics of $\{a_k,\ k\in\mathcal{M}\}$. Without loss of generality, we may assume that $a_1 \geq a_2 \geq \ldots \geq a_m$. Moreover, let $a_0 := \infty$ and $a_{m+1} := -\infty$. The objective function from (2.6) and its optimal solution, s^*, are finite due to the $(\cdot)_+$ component and the fact that a_k's are finite. For any s such that $s \in \big(a_{j+1}, a_j\big]$, where $0 \leq j \leq m$, the following are true:

- **i)** If $j = l$, then it is straightforward to see that $s + \frac{1}{l}\sum_{i\in\mathcal{M}}\big(a_i - s\big)_+ = \frac{1}{l}\sum_{i=1}^{l} a_i$;
- **ii)** If $j < l$, then we have that

$$s + \frac{1}{l}\sum_{i\in\mathcal{M}}\big(a_i - s\big)_+ = \frac{1}{l}\left(\sum_{i=1}^{j} a_i + (l-j)s\right) \geq \frac{1}{l}\left(\sum_{i=1}^{j} a_i + (l-j)a_{j+1}\right) \geq \frac{1}{l}\sum_{i=1}^{l} a_i;$$

iii) If $j > l$, then we have that

$$s + \frac{1}{l}\sum_{i\in\mathcal{M}} \big(a_i - s\big)_+ = \frac{1}{l}\left(\sum_{i=1}^{j} a_i - (j-l)s\right) \geq \frac{1}{l}\left(\sum_{i=1}^{j} a_i - (j-l)a_j\right) \geq \frac{1}{l}\sum_{i=1}^{l} a_i.$$

As a result, $s^* \in \big(a_{l+1}, a_l\big]$, which in turn justifies our claim.

The next proposition shows how to solve a weighted worst-case scenario-type optimisation problem in practice, i.e. to optimise (2.6) over $\mathbf{t} \in \mathcal{A}$, and it is given as Proposition 2.1.

Proposition 2.1. *Optimising* (2.6) *over a convex set* $\boldsymbol{t} \in \mathcal{A}$*, i.e.*

$$\min_{(\boldsymbol{t},s)\in\mathcal{A}\times\Re} \left\{ s + \frac{1}{l}\sum_{i\in\mathcal{M}} \Big(f\big(\boldsymbol{t};\boldsymbol{\omega}_i\big) - s\Big)_+ \right\},$$

is equivalent to solving

$$\min_{(\boldsymbol{t},s,\boldsymbol{u})\in\mathcal{A}\times\Re\times\Re^m} \left\{ s + \frac{1}{l}\boldsymbol{1}^T\boldsymbol{u} \right\}, \ \ s.t. \ \ \boldsymbol{0} \leq \boldsymbol{u}, \ \ f\big(\boldsymbol{t};\boldsymbol{\omega}_i\big) \leq s + u_i, \quad \forall\, i \in \mathcal{M}. \tag{2.7}$$

The computational advantage of (2.7) is conspicuous, since most of the terms are linear. Specifically, if $f\big(\cdot;\boldsymbol{\omega}_i\big)$ are SOCP representable for all $i \in \mathcal{M}$, then (2.7) becomes an SOCP problem, which could be efficiently computed.

2.2. **Optimal Robust Insurance Problem Definition.** Consider an insurance buyer who optimises its risk position by entering an insurance contract which reduces the buyer's original risk exposure $X > 0$ to $I[X]$ at a cost $P > 0$, known as the premium. Let $R[X] = X - I[X]$ denote the part of risk X ceded to the insurance seller. In order to avoid potential moral hazard issues, both I and R should be non-decreasing functions. Thus, $I, R \in \mathcal{C}^{co}$ where

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

Assume that any feasible reinsurance contract satisfies $\Phi\big(R[X];\mathcal{P}\big) \leq P \leq \bar{P}$, where $\Phi(\cdot;\mathcal{P})$ represents the *premium principle*, i.e. a certain rule of calculating the premium based on the probability measure $\mathcal{P}$. The constraint, $\Phi\big(R[X];\mathcal{P}\big) \leq P$, could be viewed as a rationality constraint. The insurance seller makes no profit before selling the insurance contract and after that, its net loss becomes $R[X] - P$. Therefore, the rationality constraint for the insurance seller becomes $\Phi\big(R[X] - P;\mathcal{P}\big) \leq 0$. The latter is equivalent to $\Phi\big(R[X];\mathcal{P}\big) \leq P$, if $\Phi(0;\mathcal{P}) = 0$ and Φ is a *translation invariant* risk measure (for details see Definition 2.1).

Definition 2.1. *Let* $(\Omega, \mathcal{F}, \mathcal{P})$ *be a probability space and* $\mathcal{X}$ *be a linear space for random variables defined on* Ω*. Then, for any* $X \in \mathcal{X}$ *and* $a \in \Re$*,* $\Phi : \mathcal{X} \to \Re$ *is a translation invariant risk measure if* $\Phi(X + a) = \Phi(X) + a$*.*

Now, when the probability measure $\mathcal{P}$ is unknown, one may be interested in finding a more robust reinsurance contract which takes into account the parameter and/or model uncertainty. Assume that there are m possible probability measures $\{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_m\}$. Then, the feasibility constraint becomes $\Phi\big(R[X]; \mathcal{P}_k\big) \leq P \leq \bar{P}$ for all $k \in \mathcal{M}$. Clearly, the insurer's net random loss is $X - R[X] + P$. Further, we assume that the insurer orders its preferences via a risk measure ρ and thus, its objective under the k^{th} model is $\rho\big(X - R[X] + P; \mathcal{P}_k\big)$, which reduces to $\rho\big(X - R[X]; \mathcal{P}_k\big) + P$ if ρ is a translation invariant risk measure.

In order to find the 'best' robust decision for the insurer, we first present four robust optimisation formulations that are detailed in Section 2.1. Their results are compared in pairs and further compared to some traditional non-robust optimal insurance arrangements. In summary, the following four robust optimisation formulations are considered for now:

A) *wc-type* as defined in (2.2)

$$\min_{(R,P)\in\mathcal{C}^{co}\times\Re} \left\{\max_{k\in\mathcal{M}} \rho\big(X - R[X]; \mathcal{P}_k\big) + P\right\} \quad \text{s.t.} \quad \Phi\big(R[X]; \mathcal{P}_k\big) \leq P \leq \bar{P} \quad \forall\, k \in \mathcal{M}. \tag{2.8}$$

B) *ad-type* as given in (2.4)

$$\min_{(R,P)\in\mathcal{C}^{co}\times\Re} \left\{\frac{1}{m}\sum_{k\in\mathcal{M}} \rho\big(X - R[X]; \mathcal{P}_k\big) + P\right\} \quad \text{s.t.} \quad \Phi\big(R[X]; \mathcal{P}_k\big) \leq P \leq \bar{P} \quad \forall\, k \in \mathcal{M}. \tag{2.9}$$

C) *wa-type* as defined in (2.4)

$$\min_{(R,P)\in\mathcal{C}^{co}\times\Re} \left\{\sum_{k\in\mathcal{M}} \lambda_k \rho\big(X - R[X]; \mathcal{P}_k\big) + P\right\} \quad \text{s.t.} \quad \Phi\big(R[X]; \mathcal{P}_k\big) \leq P \leq \bar{P} \quad \forall\, k \in \mathcal{M}. \tag{2.10}$$

D) *wwc-type* as given in (2.5)

$$\min_{(R,P)\in\mathcal{C}^{co}\times\Re} \left\{\frac{1}{l}\sum_{i=1}^{l} \rho\big(X - R[X]; \mathcal{P}_{\sigma(i)}\big) + P\right\} \quad \text{s.t.} \quad \Phi\big(R[X]; \mathcal{P}_k\big) \leq P \leq \bar{P} \quad \forall\, k \in \mathcal{M}, \tag{2.11}$$

where $\rho\big(X - R[X]; \mathcal{P}_{\sigma(i)}\big)$ is such that $\rho\big(X - R[X]; \mathcal{P}_{\sigma(1)}\big) \geq \ldots \geq \rho\big(X - R[X]; \mathcal{P}_{\sigma(l)}\big)$ with σ being a permutation of $\mathcal{M}$.

Recall that we implicitly assumed that the ρ and Φ are translation invariant risk measures, which is a very mild restriction. When $l = 1$, the $wwc - type$ Problem (2.11) becomes the $wc - type$ Problem (2.8). Moreover, when $l = m$, the $wwc - type$ Problem (2.11) becomes the $ad - type$ Problem (2.9).

3. Empirical Formulations

3.1. **Computable Formulations.** The robust optimisation problems (2.8)–(2.11) may be numerically solved by assuming a discrete distributed X with a finite sample space, i.e. the possible outcomes are $\mathbf{x} := (x_1, x_2, \ldots, x_n)^T$. Without loss of generality, one may assume that $x_1 \leq x_2 \leq \cdots \leq x_n$. The risk ceding function $R[X]$ is also discretised and becomes $\mathbf{y} := (y_1, y_2, \ldots, y_n)^T$ such that $R[X] = y_i$ if $X = x_i$ for all $1 \leq i \leq n$. Under $\mathcal{P}_k$, denote the probability vector, $\mathbf{p}_k := (p_{1k}, p_{2k}, \ldots, p_{nk})^T$, where $p_{ik} = \mathcal{P}_k(X = x_i)$ for all $1 \leq i \leq n$ and $k \in \mathcal{M}$.

Two standard risk measures used in practice that play an important role in our analysis is the *Value-at-Risk (VaR)* and *Conditional Value-at-Risk (CVaR)*. The VaR of a generic loss variable $Z > 0$ with confidence level $\alpha \in (0,1)$ is defined as

$$\mathrm{VaR}_\alpha(Z;\mathcal{P}) := \inf_{y\in\Re} \left\{\mathcal{P}(Z \leq z) \geq \alpha\right\},$$

while the CVaR is given as (see Rockafeller and Uryasev, 2000):

$$\mathrm{CVaR}_\alpha(Z;\mathcal{P}) := \inf_{t\in\Re} \left\{t + \frac{1}{1-\alpha}\mathbb{E}_{\mathcal{P}}(Z-t)_+\right\}. \tag{3.1}$$

By definition, $\mathbb{E}_{\mathcal{P}}(\cdot)$ represents the expectation with respect to $\mathcal{P}$. Note that the *wwc-type* robust risk measure given in (2.6) may be understood as a discretised version of the above representation (3.1).

Recall that $R \in \mathcal{C}^{co}$, which implies that $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{x}-\mathbf{y}$ are all non-decreasingly ordered. Therefore the empirical measure of $\mathrm{VaR}_\alpha\big(X - R[X];\mathcal{P}_k\big)$ becomes $x_{p(k)} - y_{p(k)}$, where

$$p(k) = \min_j \left\{\sum_{i=1}^{j} p_{ik} \geq \alpha\right\}.$$

On the other hand, the empirical measure of $\mathrm{CVaR}_\alpha\big(X - R[X];\mathcal{P}_k\big)$ becomes $\boldsymbol{\phi}_k^T\mathbf{x} - \boldsymbol{\phi}_k^T\mathbf{y}$, where $\boldsymbol{\phi}_k := (\phi_{1k}, \phi_{2k}, \ldots, \phi_{nk})^T$ with

$$\phi_{ik} = g\left(1 - \sum_{j=1}^{i-1} p_{jk}\right) - g\left(1 - \sum_{j=1}^{i} p_{jk}\right), \quad 1 \leq i \leq n, \;\; k \in \mathcal{M} \tag{3.2}$$

and $g(t) = \min\left(\frac{t}{1-\alpha}, 1\right)$. By convention, the summation is read as 0 when the bound of the above summation is 0.

It has been mentioned in Section 2.2 that ρ and Φ are assumed to be translation invariant risk measures in this paper. It is important to point out that we had carried out numerous numerical experiments and found that the choice of premium principle does not have an impact on our conclusions and for this reason, the numerical analysis in this paper will focus on examples with the assumption that the expected value premium principle is in force, i.e. $\Phi(\cdot;\mathcal{P}) = (1+\theta)\mathbb{E}_{\mathcal{P}}(\cdot)$ with $\theta > 0$. In turn, the premium constraints become:

$$(1+\theta)\mathbf{p}_k^T\mathbf{y} \leq P \leq \overline{P}, \quad \forall\, k \in \mathcal{M}. \tag{3.3}$$

Recall that $X > 0$ and $I, R \in \mathcal{C}^{co}$, which is equivalent to

$$\mathbf{0} \leq \mathbf{y} \leq \mathbf{x}, \mathbf{0} \leq \mathbf{A}\mathbf{y} \leq \mathbf{A}\mathbf{x}, \tag{3.4}$$

where $\mathbf{A}$ is an n-by-n matrix given by

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

We now provide the LP formulations of the robust optimisation problems (2.8)–(2.11). It is first assumed that the insurance buyer orders its preferences as via the VaR risk measure, i.e. $\rho(\cdot;\mathcal{P}) = \mathrm{VaR}_\alpha(\cdot;\mathcal{P})$. Since $X - R[X] \in \mathcal{C}^{co}$, we have that $\mathrm{VaR}_\alpha(X - R[X];\mathcal{P}_k) = x_{p(k)} - y_{p(k)}$ for all $k \in \mathcal{M}$. Therefore,

A) The *wc-type* optimisation problem from (2.8) becomes

$$\min_{(\mathbf{y},P,r)\in\Re^n\times\Re\times\Re} r \quad \text{s.t. } x_{p(k)} - y_{p(k)} + P \leq r, \quad \forall\, k \in \mathcal{M}, \quad \text{(3.3) and (3.4) hold.} \tag{3.5}$$

B) The *ad-type* optimisation problem from (2.9) becomes

$$\min_{(\mathbf{y},P)\in\Re^n\times\Re} \left\{ \frac{1}{m} \sum_{k\in\mathcal{M}} \big(x_{p(k)} - y_{p(k)}\big) + P \right\} \text{ s.t. (3.3) and (3.4) hold.} \tag{3.6}$$

C) The *wa-type* optimisation problem from (2.10) becomes

$$\min_{(\mathbf{y},P)\in\Re^n\times\Re} \left\{ \sum_{k\in\mathcal{M}} \lambda_k \big(x_{p(k)} - y_{p(k)}\big) + P \right\} \text{ s.t. (3.3) and (3.4) hold.} \tag{3.7}$$

D) The *wwc-type* optimisation problem from (2.11) becomes

$$\begin{aligned} &\min_{(\mathbf{y},P,r,s,\mathbf{u})\in\Re^n\times\Re\times\Re\times\Re\times\Re^m} \{r + P\} \\ &\text{s.t.} \quad s + \frac{1}{l}\mathbf{1}^T\mathbf{u} \leq r,\ \mathbf{0} \leq \mathbf{u},\ \text{(3.3) and (3.4) hold,} \\ &\qquad x_{p(k)} - y_{p(k)} - s \leq u_k,\ \forall\, k \in \mathcal{M}. \end{aligned} \tag{3.8}$$

The epigraph form from (3.5) is a standard reformulation in optimisation, while (3.6) and (3.7) are straightforward reformulations that do not require any additional work.

The second case is the one in which the insurance buyer orders its preferences via the CVaR risk measure, i.e. $\rho(\cdot;\mathcal{P}) = \mathrm{CVaR}_\alpha(\cdot;\mathcal{P})$. Since $X - R[X] \in \mathcal{C}^{co}$, we have that

$$\mathrm{CVaR}_\alpha(X - R[X];\mathcal{P}_k) = \boldsymbol{\phi}_k^T\mathbf{x} - \boldsymbol{\phi}_k^T\mathbf{y}, \quad \forall\, k \in \mathcal{M},$$

by keeping in mind (3.2). Therefore, (2.8)–(2.11) are equivalent to solving the following optimisation problems:

A) The *wc-type* optimisation problem from (2.8) becomes

$$\min_{(\mathbf{y},P,r)\in\Re^n\times\Re\times\Re} r \quad \text{s.t. } \boldsymbol{\phi}_k^T\mathbf{x}-\boldsymbol{\phi}_k^T\mathbf{y}+P\le r, \quad \forall\, k\in\mathcal{M}, \quad \text{(3.3) and (3.4) hold.} \tag{3.9}$$

B) The *ad-type* optimisation problem from (2.9) becomes

$$\min_{(\mathbf{y},P)\in\Re^n\times\Re} \left\{\frac{1}{m}\sum_{k\in\mathcal{M}}\left(\boldsymbol{\phi}_k^T\mathbf{x}-\boldsymbol{\phi}_k^T\mathbf{y}\right)+P\right\} \text{ s.t. (3.3) and (3.4) hold.} \tag{3.10}$$

C) The *wa-type* optimisation problem from (2.10) becomes

$$\min_{(\mathbf{y},P)\in\Re^n\times\Re} \left\{\sum_{k\in\mathcal{M}}\lambda_k\left(\boldsymbol{\phi}_k^T\mathbf{x}-\boldsymbol{\phi}_k^T\mathbf{y}\right)+P\right\} \text{ s.t. (3.3) and (3.4) hold.} \tag{3.11}$$

D) The *wwc-type* optimisation problem from (2.11) becomes

$$\begin{aligned}&\min_{(\mathbf{y},P,r,s,\mathbf{u})\in\Re^n\times\Re\times\Re\times\Re\times\Re^m} \{r+P\} \\ &\text{s.t.}\quad s+\frac{1}{l}\mathbf{1}^T\mathbf{u}\le r,\ \mathbf{0}\le\mathbf{u},\ \text{(3.3) and (3.4) hold,} \\ &\qquad \boldsymbol{\phi}_k^T\mathbf{x}-\boldsymbol{\phi}_k^T\mathbf{y}-s\le u_k,\ \forall\ k\in\mathcal{M}.\end{aligned} \tag{3.12}$$

The epigraph form from (3.9) is a standard reformulation in optimisation, while (3.10) and (3.11) are straightforward reformulations that do not require any additional work.

Next, we assume that the insurance buyer orders its preferences as via the PHT risk measure, i.e. $\rho(\cdot;\mathcal{P})=\int_0^\infty g\big(\mathcal{P}(\cdot>x)\big)\,dx$ with $g(t)=t^\alpha$, $0<\alpha\le 1$ (for details, see Wang *et al.*, 1997). Since $X-R[X]\in\mathcal{C}^{co}$, we have that

$$PHT_\alpha(X-R[X];\mathcal{P}_k)=\boldsymbol{\phi}_k^T\mathbf{x}-\boldsymbol{\phi}_k^T\mathbf{y}, \quad \forall\ k\in\mathcal{M},$$

where $\boldsymbol{\phi}_k$ are defined as in (3.2) with $g(t)=t^\alpha$. Therefore, the robust optimisation problems (2.8)–(2.11) are precisely as in (3.9)–(3.12), but with different parameters $\boldsymbol{\phi}_k$'s.

The final case is when the insurance buyer orders its risk preferences as via the *standard deviation SD* risk measure, i.e. $\rho(\cdot;\mathcal{P})=E_{\mathcal{P}}(\cdot)+bSd(\cdot;\mathcal{P})$ with $b>0$. For a generic discrete random variable Z with a finite sample space $(z_1,z_2,\ldots,z_n)$ that is equipped with a probability measure $\mathcal{P}$ such that $\mathcal{P}(Z=z_j)=p_j$, its standard deviation can be written as $Sd(Z;\mathcal{P})=\|\mathbf{Qz}\|$, where $\mathbf{Q}$ is a $n\times n$ matrix with its (j_1,j_2)-th element to be $q_{j_1j_2}=\sqrt{p_{j_1}}(1_{j_1=j_2}-p_{j_2})$ for all $1\le j_1,j_2\le n$. By definition, 1_A represent the indicator operator and takes the value one if A is true and to take the value zero otherwise. Therefore, the SD risk measure under $\mathcal{P}_k$ can be written as

$$\rho\big(X-R[X];\mathcal{P}_k\big)=\mathbf{p}_k^T(\mathbf{x}-\mathbf{y})+b\|\mathbf{Q}_k(\mathbf{x}-\mathbf{y})\|,$$

where $q_{j_1j_2k}=\sqrt{p_{j_1k}}(1_{j_1=j_2}-p_{j_2k})$ for all $1\le j_1,j_2\le n$ and $k\in\mathcal{M}$. Note that the corresponding formulations are in SOCP form as follows:

A) The *wc-type* optimisation problem from (2.8) becomes

$$\min_{(\mathbf{y},P,r)\in\Re^n\times\Re\times\Re} r \quad \text{s.t. } \mathbf{p}_k^T(\mathbf{x}-\mathbf{y})+b\|\mathbf{Q}_k(\mathbf{x}-\mathbf{y})\|+P \le r,\ \forall\, k\in \mathcal{M},\ (3.3) \text{ and } (3.4) \text{ hold.} \tag{3.13}$$

B) The *ad-type* optimisation problem from (2.9) becomes

$$\begin{aligned} &\min_{(\mathbf{y},P,\mathbf{t})\in\Re^n\times\Re\times\Re^m} \left\{ \frac{1}{m}\sum_{k\in\mathcal{M}} \left(\mathbf{p}_k^T(\mathbf{x}-\mathbf{y})+bt_k\right)+P \right\} \\ &\text{s.t.}\quad \|\mathbf{Q}_k(\mathbf{x}-\mathbf{y})\| \le t_k,\ \forall\, k\in\mathcal{M},\ (3.3) \text{ and } (3.4) \text{ hold.} \end{aligned} \tag{3.14}$$

C) The *wa-type* optimisation problem from (2.10) becomes

$$\begin{aligned} &\min_{(\mathbf{y},P,\mathbf{t})\in\Re^n\times\Re\times\Re^m} \left\{ \sum_{k\in\mathcal{M}} \lambda_k \left(\mathbf{p}_k^T(\mathbf{x}-\mathbf{y})+bt_k\right)+P \right\} \\ &\text{s.t.}\quad \|\mathbf{Q}_k(\mathbf{x}-\mathbf{y})\| \le t_k,\ \forall\, k\in\mathcal{M},\ (3.3) \text{ and } (3.4) \text{ hold.} \end{aligned} \tag{3.15}$$

D) The *wwc-type* optimisation problem from (2.11) becomes

$$\begin{aligned} &\min_{(\mathbf{y},P,r,s,\mathbf{u})\in\Re^n\times\Re\times\Re\times\Re\times\Re^m} \{r+P\} \\ &\text{s.t.}\quad s+\frac{1}{l}\mathbf{1}^T\mathbf{u}\le r,\ \mathbf{0}\le\mathbf{u},\ (3.3) \text{ and } (3.4) \text{ hold,} \\ &\qquad \mathbf{p}_k^T(\mathbf{x}-\mathbf{y})+b\|\mathbf{Q}_k(\mathbf{x}-\mathbf{y})\|-s\le u_k,\ \forall\, k\in\mathcal{M}. \end{aligned} \tag{3.16}$$

3.2. **Pareto Optimality.** One major concern regarding the robust optimisation models from (2.8)–(2.11) is that optimal solutions could be inefficient insurance contracts. In other words, the resulting robust optimal solutions are not necessarily Pareto optimal. The idea of Pareto optimality ensures that the allocated risk is shared in the most efficient way, i.e. there is no alternative allocation that may put the insurance players in a "better" risk position. The mathematical formulation of this definition is now given in our context. That is, a robust optimal solution (R^*, P^*) is also Pareto optimal if and only if there exists no other feasible solution $(\tilde{R}, \tilde{P})$ such that

$$\rho\big(X-\tilde{R}[X];\mathcal{P}_k\big)+\tilde{P}\le\rho\big(X-R^*[X];\mathcal{P}_k\big)+P^* \qquad \forall\, k\in\mathcal{M},$$

with at least one inequality sign being strict. It is well-known that if all weighting coefficients from (2.10) are strictly positive, then its robust optimal solutions (R^*, P^*) are also Pareto optimal. That is, the solutions of the Additive Model (2.9) and Weighted Average Model (2.10) with strictly positive λ_k's (for all $k\in\mathcal{M}$) are Pareto optimal. Unfortunately, the solutions of (2.8) may lead to solutions that are not Pareto optimal, but a remedy is possible (for details, see Asimit *et al.*, 2017). The same conclusion is drawn for the solutions of (2.11) when $l<m$ and we would like to check which solutions of (2.11) are Pareto optimal and if possible, to modify those solutions of (2.11) that are not Pareto optimal into Pareto optimal solutions that solve

(2.11). This would be a generalisation of Theorem 5.1 in Asimit *et al.* (2017), which in fact is possible and we state this result as Theorem 3.1. Before giving the main result of the section, let us explain the general setting of Theorem 3.1, which is given in Problem 3.1. Note that all of the example from before have shown to be particular cases of Problem 3.1.

Problem 3.1. *Let $f_k : \mathcal{A} \to \Re, g_k : \mathcal{A} \to \Re^{n_k}$ be some functions over a convex set $\mathcal{A}$, where n_k are some positive integers, for all $k \in \mathcal{M}$. Moreover, l is an integer such that $0 < l \leq m$. Let $f^{(i)}(\cdot)$ be the i^{th} upper order statistics of $\{f_k(\cdot), k \in \mathcal{M}\}$, i.e.*

$$f^{(i)}(\cdot) = f_{\sigma(i)}(\cdot) \quad \textit{such that} \quad f_{\sigma(1)}(\cdot) \geq f_{\sigma(2)}(\cdot) \geq \ldots f_{\sigma(m)}(\cdot)$$

with σ being a permutation of $\mathcal{M}$. The optimisation problem becomes:

$$\min_{\boldsymbol{x} \in \mathcal{A}} \sum_{i=1}^{l} \lambda_i f^{(i)}(\boldsymbol{x}), \qquad s.t \;\; g_k(\boldsymbol{x}) \in \mathcal{A}_k, \;\; \forall \; k \in \mathcal{M}, \tag{3.17}$$

where λ_k's are positive scalars and $\mathcal{A}_k$ are convex cones[2] for all $k \in \mathcal{M}$.

Recall that Problem 3.1 is convex as long as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_l$ and all functions f_k, g_k are convex over $\mathcal{A}$. Using the notation from Problem 3.1, a feasible solution $\mathbf{x}^*$, i.e. $g_k(\mathbf{x}^*) \in \mathcal{A}_k$ for all $k \in \mathcal{M}$, is Pareto optimal if there is no other feasible solution $\mathbf{y}$, i.e. $g_k(\mathbf{y}) \in \mathcal{A}_k$ for all $k \in \mathcal{M}$, such that $f_k(\mathbf{y}) \leq f_k(\mathbf{x}^*)$ for all $k \in \mathcal{M}$ with at least one inequality sign being strict. We are now ready to state the main result of this section, which shows that one may identify the group of solutions of (3.17) that are Pareto optimal as well without massively increasing the computational effort.

Theorem 3.1. *Let $\boldsymbol{x}^*$ be an optimal solution of* (3.17)*. Then, $\boldsymbol{x}^*$ is also Pareto optimal if the optimal objective function value of the following optimisation problem*

$$\min_{\boldsymbol{y} \in \mathcal{A}} \sum_{k \in \mathcal{M}} \big(f_k(\boldsymbol{y}) - f_k(\boldsymbol{x}^*)\big), \quad s.t. \;\; g_k(\boldsymbol{y}) \in \boldsymbol{A}_k, \;\; f_k(\boldsymbol{y}) - f_k(\boldsymbol{x}^*) \leq 0, \;\; \forall \; k \in \mathcal{M} \tag{3.18}$$

is zero. On the other hand, if the optimal value of (3.18) *is negative, then any optimal solution $\boldsymbol{y}^*$ of* (3.18) *solves* (3.17) *as well and is Pareto optimal.*

4. Numerical results

The current section provides numerical illustrations to the robust optimisation problems 2.8–2.11. Recall that our empirical method requires a sample of $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ to be drawn from the underlying distribution of X. Note that the empirical formulations discussed in Section 3.1 are not restricted to certain distributions of X. Thus, without loss of generality, we

[2] A set $\mathcal{B}$ is a convex cone if and only if for any scalars $a, b > 0$, $a\mathbf{x} + b\mathbf{y} \in \mathcal{B}$ given that $\mathbf{x}, \mathbf{y} \in \mathcal{B}$.

further assume that X is Log-Normal distributed with mean $E(X) = 5,000$ and standard deviation $\sqrt{3} \times E(X)$. The Log-Normal assumption is one of the most natural parametric choices, as it covers distributions with very different tail distributions, i.e. from moderately light tailed to moderately heavy tailed distributions. The practitioners' literature is often based on the Log-Normal risk distribution assumption, e.g. Solvency II recommendations (see QIS 5) heavily rely on this assumption. The premium principle Φ is assumed to be an expected value principle with a risk loading factor $\theta = 0.25$, i.e. $\Phi(\cdot;\mathcal{P}) = (1+\theta)E_{\mathcal{P}}(\cdot)$. Also, the upper boundary of the maximum acceptable insurance cost is $\overline{P} = \frac{(1+\theta)E(X)}{2}$. Furthermore, the following five models are considered as potential candidates for the unknown underlying distribution of X:

(i) Model 1: Exponential distribution with mean $1/\nu$;
(ii) Model 2: Log-Normal distribution with parameters $\left(\mu, \sigma^2\right)$;
(iii) Model 3: Pareto distribution with parameters (α, λ) and cdf $F(z) = 1 - \left(\frac{\lambda}{\lambda+z}\right)^{\alpha}, z > 0$;
(iv) Model 4: Weibull distribution with parameters (c, γ) and cdf $F(z) = 1 - e^{-cz^{\gamma}}, z > 0$;
(v) Model 5: Inverse Gaussian distribution with parameters (μ, σ) and cdf
$F(z) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right) + \Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right)e^{2\lambda/\mu}, z > 0.$

For implementation purposes, we should define the probability vector $\mathbf{p}_k$ for all $k \in \mathcal{M}$ by discritising the Maximum Likelihood estimated model with the sample observation $\mathbf{x}$. That is,

$$p_{ik} = F_k\left(\frac{x_{i+1} + x_i}{2}; \hat{\nu}\right) - F_k\left(\frac{x_i + x_{i-1}}{2}; \hat{\nu}\right), \quad \text{for all} \;\; i = 1, \ldots, n, \;\; k \in \{1,2,3,4,5\}, \tag{4.1}$$

where by convention $x_0 = -\infty$ and $x_{n+1} = \infty$. Moreover, $\hat{\nu}$ is the Maximum Likelihood Estimate based on the sample $\mathbf{x}$. Let us also denote the true underlying distribution of X and its corresponding probability vector as Model 0 and $\mathbf{p}_0$, respectively. Then, $\mathbf{p}_0$ can be found by applying the formula (4.1) with $\hat{\nu}$ replaced by the Model 0 parameters. It would be interesting to see how the performance of our numerical results would be affected by the decision-maker's information set regarding the underlying distribution of X. That is, we repeat the numerical experiments for different model collections. In particular, we choose the following uncertainty sets: $\mathcal{M}_5 := \{1,2,3,4,5\}$, $\mathcal{M}_4 := \{1,3,4,5\}$, $\mathcal{M}_2 := \{1,5\}$, $\mathcal{M}_4^* := \{2,3,4,5\}$ and $\mathcal{M}_2^* := \{2,5\}$. Note that the underlying distribution of X is Log-Normal, and thus, we have deliberately excluded Model 2 from $\mathcal{M}_2$ and $\mathcal{M}_4$ in order to investigate the impact of model misidentification, when the "true" model is discarded.

We also need to specify the weights λ_k's that appear in (2.10). This is done by using the *relative likelihood (RL)* and $RL_k := e^{(AIC_{min} - AIC_k)/2}$, where $AIC_k = 2q_k - 2Ln(\hat{L}_k)$ with q_k being the number of parameters estimated under the k^{th} candidate distribution and $\hat{L}_k$ being the corresponding maximum likelihood function value. Moreover, $AIC_{min} := \min_{k\in\mathcal{M}} AIC_k$.

Finally, the weights are defined as follows: $\lambda_k := \frac{RL_k}{\sum_{k=1}^{m} RL_k}$. Note that

$$0 < RL_k \leq 1, \quad 0 < \lambda_k < 1 \quad \text{and} \quad \sum_{k=1}^{m} \lambda_k = 1 \quad \text{for all } k \in \mathcal{M}.$$

Also, if we denote k^* such that $AIC_{min} = AIC_{k^*}$, then $RL_{k^*} = 1$ and $\lambda_{k^*} \geq \lambda_k$ for all $k \in \mathcal{M}$. In other words, the "best" model based on the AIC criterion receives the largest weight.

Let us denote the optimal solutions to the robust optimisation problems (2.8)–(2.11) as $(\mathbf{y}^*_{wc}, P^*_{wc})$, $(\mathbf{y}^*_{ad}, P^*_{ad})$, $(\mathbf{y}^*_{wa}, P^*_{wa})$ and $(\mathbf{y}^*_{wwc}, P^*_{wwc})$, respectively. In particular, $\mathbf{y}^*_r$ represent the optimal insurance contract and is an n-dimensional column vector with $r \in \{wc, ad, wa, wwc\}$, while P^*_r represents the optimal insurance price and is a scalar. In order to assess the quality of our robust solutions, it is necessary to set a benchmark; a natural and fair choice is the optimal insurance contract if the underlying distribution of X would have been known, denoted by $(\mathbf{y}^*_T, P^*_T)$. In fact, $(\mathbf{y}^*_T, P^*_T)$ could be obtained by solving (2.9) with $\mathcal{M} = \{0\}$. The robustness of a generic optimal solution $\mathbf{y}^*$ is our main focus, and therefore, we could compare various optimal solutions via the following absolute error:

$$\Delta^* = \sum_{i=1}^{n} |y^*_i - y^*_{iT}| \times p_{i0}.$$

Specifically, given two optimal solutions $\mathbf{y}^*_A$ and $\mathbf{y}^*_B$, model A is preferred if $\Delta^*_A < \Delta^*_B$ and we write $S^*_A \succ S^*_B$.

The "robust" optimal solutions are compared with two "non-robust" optimal solutions. The first "non-robust" model chooses the "best" distribution for X via the *Akaike Information Criterion (AIC)*, and hence, the model is called the *AIC Model* and its optimal solution is denoted as $(\mathbf{y}^*_{AIC}, P^*_{AIC})$. The second "non-robust" model is called the *Elicitable Model* and its solution is denoted as $(\mathbf{y}^*_e, P^*_e)$. Before presenting our results, we first provide brief explanations regarding the construction of the AIC and Elicitable Models. The AIC model chooses the 'best' distribution for X among all candidate distributions by finding the distribution k which gives the smallest AIC value, i.e. $k^* := \arg\min_{k \in \mathcal{M}} AIC_k$. Then, $(\mathbf{y}^*_{AIC}, P^*_{AIC})$ is found by solving (2.9) with $\mathcal{M} = k^*$.

We now move to the construction of the Elicitable Model starting with explaining the elicitability concept. By definition, a scoring function $S : \Re \times \Re \to [0, \infty)$ is a mapping $(u, v) \mapsto S(u, v)$, where u is a point forecast and v is an observation.

Definition 4.1. *Let $f : \Pi \to 2^{\Re}$ be a functional on a class of probability measures Π on $\Re$ such that $\mathcal{P} \mapsto f(\mathcal{P}) \subset \Re$, where $\mathcal{P} \in \Pi$. A scoring function $S : \Re \times \Re \to [0, \infty)$ is consistent for the functional f relative to Π if and only if $\mathbb{E}_{\mathcal{P}} S(t, L) \leq \mathbb{E}_{\mathcal{P}} S(z, L)$ for all $\mathcal{P} \in \Pi, t \in f(\mathcal{P})$ and*

$x \in \Re$. *Moreover, S is a strictly consistent scoring function if S is consistent and*

$$\mathbb{E}_{\mathcal{P}} S(t, L) = \mathbb{E}_{\mathcal{P}} S(z, L) \implies z \in f(\mathcal{P}).$$

The functional f is *elicitable* relative to a class of probability measure Π if and only if there exists a scoring function S that is strictly consistent for f relative to Π. The concept of elicitable risk measure is introduced by Lambert *et. al.* (2008), but a comprehensive background about elicitability could be found in the seminal paper of Gneiting (2011). The latter paper tells us that VaR_α is elicitable and

$$\mathbb{E}_{\mathcal{P}} S_g\big(\text{VaR}_\alpha(X; \mathcal{P}), x\big) \le \mathbb{E}_{\mathcal{P}} S_g\big(y, x\big) \tag{4.2}$$

for any real number $y \in \Re$, where $S_g(t, x) = \big(\mathrm{I}_{\{t \ge x\}} - \alpha\big)\big(g(t) - g(x)\big)$ is the scoring function and g is any non-decreasing function. The translation of (4.2) into our discretised empirical formulation under any probability distribution $\mathcal{P}_k$ becomes

$$\sum_{i=1}^{n} p_{ik} S_g\big(x_{p(k)}, x_i\big) \le \sum_{i=1}^{n} p_{ik} S_g\big(y, x_i\big).$$

As a result, whenever the "true" probability distribution $\mathcal{P}_k$ is unknown, but m probability candidate models are available, one may choose the "best" distribution k^* that gives the lowest expected score, i.e.

$$k^* = \arg\min_k \sum_{i=1}^{n} p_{ik} S\left(x_{p(k)}, x_i\right),$$

and hence, the "best" estimate of VaR_α is $x_{p(k^*)}$. Finally, the non-robust optimal elicitability solution $(\mathbf{y}_e^*, P_e^*)$ may be found by solving the following LP for all $l \in \mathcal{M}$:

$$\begin{aligned} \min_{(\mathbf{y}, P) \in \Re^n \times \Re} \quad & \left\{x_{p(l)} - y_{p(l)} + P\right\} \\ \text{s.t.} \quad & \sum_{i=1}^{n} p_{il} S\big(x_{p(l)} - y_{p(l)}, x_i - y_i\big) \le \sum_{i=1}^{n} p_{il} S\big(x_{p(l)} - y_{p(l)}, x_i - y_i\big), \ \forall\, k \in \mathcal{M}, \\ & (1+\theta)\mathbf{p}_k^T \mathbf{y} \le P \le \overline{P}, \ \forall\, k \in \mathcal{M}, \\ & \mathbf{0} \le \mathbf{y} \le \mathbf{x}, \mathbf{0} \le \mathbf{A}\mathbf{y} \le \mathbf{A}\mathbf{x}. \end{aligned} \tag{4.3}$$

Let $(\mathbf{y}_{el}^*, P_{el}^*)$ be the optimal solution found for the above LP under distribution l and let l^* be the probability model choice under the elicitability criterion, which is given by the one with the lowest objective function (4.3) amongst all $l \in \mathcal{M}$. Therefore, the Elicitability Model optimal solution is $(\mathbf{y}_e^*, P_e^*) := (\mathbf{y}_{el^*}^*, P_{el^*}^*)$. Recall that all other risk measures considered in this paper, i.e. CVaR, PHT and SD are not elicitable, although CVaR and VaR are jointly elicitable, and therefore, the Elicitable Model is only applied with the VaR-based case.

Before discussing the results of our numerical experiments, we note that all optimisation problems are implemented on a desktop with 6 core Intel i7-5820K at 3.30GHz, 16GB RAM, running Linux x64, MATLAB R2014b, CVX 2.1.

4.1. **Comparison of Robustness.** We first investigate the results for the VaR_α-based optimisation problems when $\alpha = 0.75$, which are illustrated in Tables 4.1–4.3. Our numerical

	$n = 25$			$n = 50$			$n = 100$			$n = 250$		
	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$
$S^*_{wc} \succ S^*_{wa}$	215	224	226	184	187	213	126	125	159	77	76	131
$S^*_{wa} \succ S^*_{wc}$	285	276	273	316	313	287	374	375	341	423	424	369
$S^*_{ad} \succ S^*_{wa}$	234	204	207	225	218	237	171	140	186	126	107	141
$S^*_{wa} \succ S^*_{ad}$	219	243	237	275	279	260	329	354	314	374	393	358
$S^*_{wa} \succ S^*_{AIC}$	224	216	221	198	198	211	127	164	152	56	171	71
$S^*_{AIC} \succ S^*_{wa}$	275	284	278	302	302	289	373	336	347	444	329	429
$S^*_{wa} \succ S^*_{e}$	320	352	349	386	424	417	457	469	455	494	494	489
$S^*_{e} \succ S^*_{wa}$	180	148	149	114	76	83	43	31	45	6	6	11

TABLE 4.1. Results when (3.7) is compared to (3.5), (3.6) and the AIC model for the $\mathrm{VaR}_{0.75}$-based solutions under various sample sizes n and collections of candidate models $\{\mathcal{M}_2, \mathcal{M}_4, \mathcal{M}_5\}$.

experiments are set for 500 samples of various sizes $n = \{25, 50, 100, 250\}$ and results are reported as the number of experiments out of 500 in which a particular model is preferred when compared to another. The top four rows in Tables 4.1 and 4.2 together with Table 4.3 show the results when the Weighted Average Model (3.7) is compared to the other three robust models (3.5), (3.6) and (3.8), respectively. Recall that when $l = 1$, the Weighted Worst-case Model (3.8) becomes the classic Worst-case model (3.5), and thus, we only solve (3.8) under $\mathcal{M} = \mathcal{M}_5, \mathcal{M}_4$ and $\mathcal{M}_4^*$ with $2 \leq l \leq m-1$. Note that the $l = 4$ case only exists when $\mathcal{M} = \mathcal{M}_5$. We noticed that the Weighted Average Model stands as the most robust model in all comparisons, especially when the true underlying distribution of X is not included in the candidate distribution collection $\mathcal{M}$, i.e. under $\mathcal{M}_4$ and $\mathcal{M}_2$. The last four rows from Tables 4.1 and 4.2 compare the solutions found under the Weighted Average Model (3.7) to those found under the non-robust models, i.e. the AIC and the Elicitable Models. It is surprisingly clear that the Weighted Average Model (3.7) does outperform the elicitability criterion. However, the performance of the

	$n=25$			$n=50$			$n=100$			$n=250$		
	$\mathcal{M}_5$	$\mathcal{M}_4^*$	$\mathcal{M}_2^*$	$\mathcal{M}_5$	$\mathcal{M}_4^*$	$\mathcal{M}_2^*$	$\mathcal{M}_5$	$\mathcal{M}_4^*$	$\mathcal{M}_2^*$	$\mathcal{M}_5$	$\mathcal{M}_4^*$	$\mathcal{M}_2^*$
$S^*_{wc} \succ S^*_{wa}$	215	211	246	184	162	259	126	108	236	77	40	264
$S^*_{wa} \succ S^*_{wc}$	285	289	254	316	338	240	374	392	264	423	460	234
$S^*_{ad} \succ S^*_{wa}$	234	227	130	225	240	160	171	203	196	126	150	219
$S^*_{wa} \succ S^*_{ad}$	219	211	131	275	255	166	329	297	223	374	350	245
$S^*_{wa} \succ S^*_{AIC}$	224	177	236	198	170	218	127	134	248	56	107	272
$S^*_{AIC} \succ S^*_{wa}$	275	323	263	302	330	281	373	366	252	444	393	224
$S^*_{wa} \succ S^*_{e}$	320	291	275	386	354	281	457	406	295	494	455	271
$S^*_{e} \succ S^*_{wa}$	180	209	223	114	146	218	43	93	204	6	45	224

TABLE 4.2. Results when (3.7) is compared to (3.5), (3.6) and the AIC model for the $\text{VaR}_{0.75}$-based solutions under various sample sizes n and collections of candidate models $\{\mathcal{M}_2^*, \mathcal{M}_4^*, \mathcal{M}_5\}$.

	$n=25$			$n=50$			$n=100$			$n=250$		
	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$
$S^*_{wwc} \succ S^*_{wa}\ (l=2)$	202	201	204	171	182	160	123	116	106	77	78	40
$S^*_{wa} \succ S^*_{wwc}\ (l=2)$	298	299	296	329	317	340	377	384	394	423	422	460
$S^*_{wwc} \succ S^*_{wa}\ (l=3)$	238	227	259	182	202	219	128	136	163	84	81	109
$S^*_{wa} \succ S^*_{wwc}\ (l=3)$	262	273	241	318	298	281	372	364	337	416	419	391
$S^*_{wwc} \succ S^*_{wa}\ (l=4)$	241			203			153			105		
$S^*_{wa} \succ S^*_{wwc}\ (l=4)$	259			297			347			395		

TABLE 4.3. Comparison between the $\text{VaR}_{0.75}$-based solutions of (3.7) and (3.8) for various sample sizes n and collections of candidate models $\{\mathcal{M}_5, \mathcal{M}_4, \mathcal{M}_4^*\}$.

robust models are uniformly weaker than the non-robust AIC model across various combinations of sample sizes and distribution collections. Similar outcomes may be found in Asimit *et al.* (2017), where it is argued that such peculiar behaviour is due to the robustness of VaR itself as a risk measure.

Recall that the comparison between the optimal contracts is done by looking into the Δ^* values, but these may be misleading if these values are quite small. Thus, additional comparisons would help in getting more confidence in our results and boxplots of Δ^*'s might be informative as well. Figure 4.1 compares the boxplots between Δ^*_{wa} and Δ^*_{AIC}. In each of the boxplots,

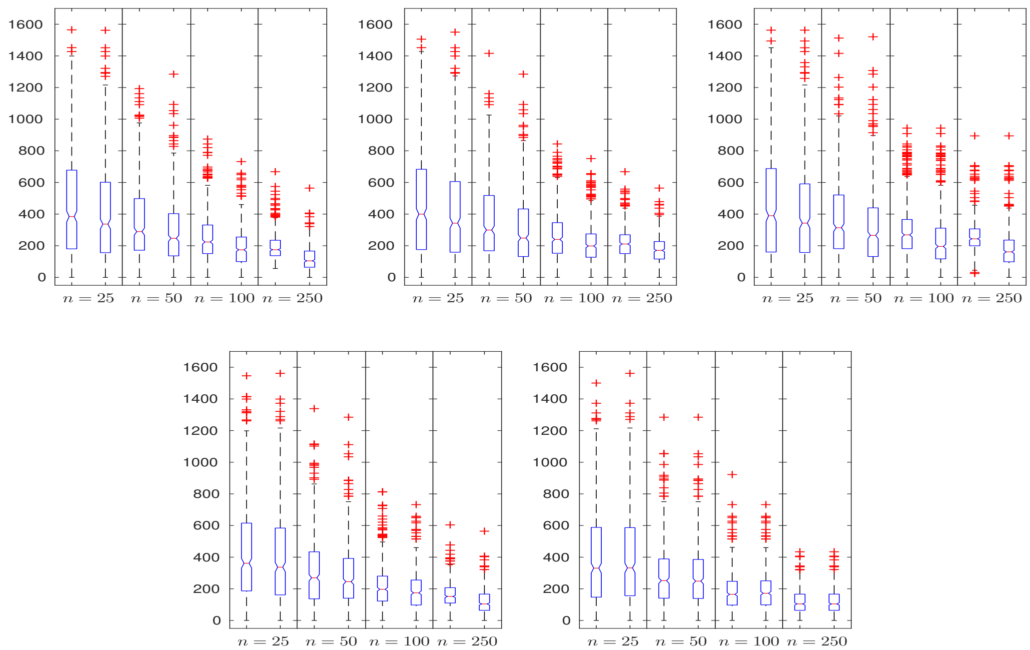

FIGURE 4.1. Boxplots comparing Δ^*_{wa} and Δ^*_{AIC} computed from the $\text{VaR}_{0.75}$-based optimisation cases. Each graph constitutes of four groups of boxplots that correspond to various sample sizes of n. The boxplot on the left/right-hand side represents $\Delta^*_{wa}/\Delta^*_{AIC}$. The top row boxplots are corresponding to distribution collections $\mathcal{M}_5$, $\mathcal{M}_4$ and $\mathcal{M}_2$, while the bottom row relates to $\mathcal{M}^*_4$ and $\mathcal{M}^*_2$, respectively.

the median of Δ^*'s is marked by a short red line inside the notched box, while the box itself represents the inter-quartile range. All outliers are marked by a red cross. It is not difficult to see that the variation of both Δ^*_{wa} and Δ^*_{AIC} shrinks dramatically when the sample size n grows for all distribution collections $\mathcal{M} \in \{\mathcal{M}_5, \mathcal{M}_4, \mathcal{M}_2, \mathcal{M}^*_4, \mathcal{M}^*_2\}$. It is also worth pointing out that although Tables 4.1 and 4.2 tell us that the AIC Model is preferred to all robust optimisation models (3.5)–(3.11) in the VaR-based case, Figure 4.1 shows that Δ^*_{wa} and Δ^*_{AIC} have quite similar ranges, especially when the sample size n is small.

Next, we turn our attention to the set of results relating to the $\text{CVaR}_{0.75}$-based decisions which are given in Tables 4.4–4.6. Similar to the VaR-based case, we first compare among the robust optimal solutions found in (3.9)–(3.12). Tables 4.4–4.6 have shown a similar pattern as seen in the VaR case, where the optimal solutions found under the Weighted Average Model (3.11) turn out to be the most robust among the four models (3.9)–(3.12), especially when n is large. Further, there is strong numerical evidence showing that the Weighted Average Model performs uniformly better than the non-robust AIC model throughout various combinations of sample

	$n=25$			$n=50$			$n=100$			$n=250$		
	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$
$S^*_{wc} \succ S^*_{wa}$	189	231	203	198	194	191	180	172	225	155	210	281
$S^*_{wa} \succ S^*_{wc}$	311	269	297	300	306	307	320	328	275	345	290	219
$S^*_{ad} \succ S^*_{wa}$	259	263	278	252	272	285	244	254	279	128	207	221
$S^*_{wa} \succ S^*_{ad}$	240	237	221	248	227	213	256	246	221	372	293	279
$S^*_{wa} \succ S^*_{AIC}$	276	261	269	298	301	294	271	285	267	253	213	250
$S^*_{AIC} \succ S^*_{wa}$	224	239	231	202	199	206	229	215	233	247	287	250

TABLE 4.4. Results when (3.11) is compared to (3.9), (3.10) and the AIC model for the $\text{CVaR}_{0.75}$-based solutions under various sample sizes n and collections of candidate models $\{\mathcal{M}_2, \mathcal{M}_4, \mathcal{M}_5\}$.

	$n=25$			$n=50$			$n=100$			$n=250$		
	$\mathcal{M}_5$	$\mathcal{M}^*_4$	$\mathcal{M}^*_2$	$\mathcal{M}_5$	$\mathcal{M}^*_4$	$\mathcal{M}^*_2$	$\mathcal{M}_5$	$\mathcal{M}^*_4$	$\mathcal{M}^*_2$	$\mathcal{M}_5$	$\mathcal{M}^*_4$	$\mathcal{M}^*_2$
$S^*_{wc} \succ S^*_{wa}$	189	197	221	198	180	192	180	183	188	155	166	202
$S^*_{wa} \succ S^*_{wc}$	311	303	279	300	320	308	320	317	312	345	334	298
$S^*_{ad} \succ S^*_{wa}$	259	258	268	252	262	243	244	274	231	128	168	242
$S^*_{wa} \succ S^*_{ad}$	240	238	232	248	236	254	256	226	267	372	332	258
$S^*_{wa} \succ S^*_{AIC}$	276	281	286	298	301	282	271	269	272	253	255	248
$S^*_{AIC} \succ S^*_{wa}$	224	219	213	202	199	218	229	131	227	247	245	252

TABLE 4.5. Results when (3.11) is compared to (3.9), (3.10) and the AIC model for the $\text{CVaR}_{0.75}$-based solutions under various sample sizes n and collections of candidate models $\{\mathcal{M}^*_2, \mathcal{M}^*_4, \mathcal{M}_5\}$.

sizes n and distribution collections M. Boxplots are also produced to better compare Δ^*_{wa} and Δ_{AIC} for CVaR-based optimisations, which could be found in Figure 4.2. Although the median value of Δ^*_{wa} and Δ^*_{AIC} are very similar under various sample sizes and distribution collections, the range of Δ^*_{AIC} is in general larger than that of Δ^*_{wa}, especially when the sample is small. Therefore, the overall evidence tells us that our Weighted Average Model (3.11) leads to the most robust optimal solution for CVaR-based decisions.

The third set of results are related to the PHT-based optimal solutions from (3.9)–(3.12) and the AIC model. The results from Tables 4.7–4.9 tell us that the Weighted Average Model performs better than all other "robust" models, which is even more evident when the sample size

	$n = 25$			$n = 50$			$n = 100$			$n = 250$		
	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$
$S^*_{wwc} \succ S^*_{wa}$ $(l=2)$	199	268	203	204	262	203	208	277	197	195	321	203
$S^*_{wa} \succ S^*_{wwc}$ $(l=2)$	301	232	297	296	238	297	292	223	303	305	179	297
$S^*_{wwc} \succ S^*_{wa}$ $(l=3)$	233	243	248	225	258	229	250	264	259	251	182	239
$S^*_{wa} \succ S^*_{wwc}$ $(l=3)$	266	257	252	274	242	271	250	236	241	248	318	261
$S^*_{wwc} \succ S^*_{wa}$ $(l=4)$	235			242			253			156		
$S^*_{wa} \succ S^*_{wwc}$ $(l=4)$	265			258			247			344		

TABLE 4.6. Comparison between the $\mathrm{CVaR}_{0.75}$-based solutions of (3.11) and (3.12) for various sample sizes n and collections of candidate models $\{\mathcal{M}_5, \mathcal{M}_4, \mathcal{M}_4^*\}$.

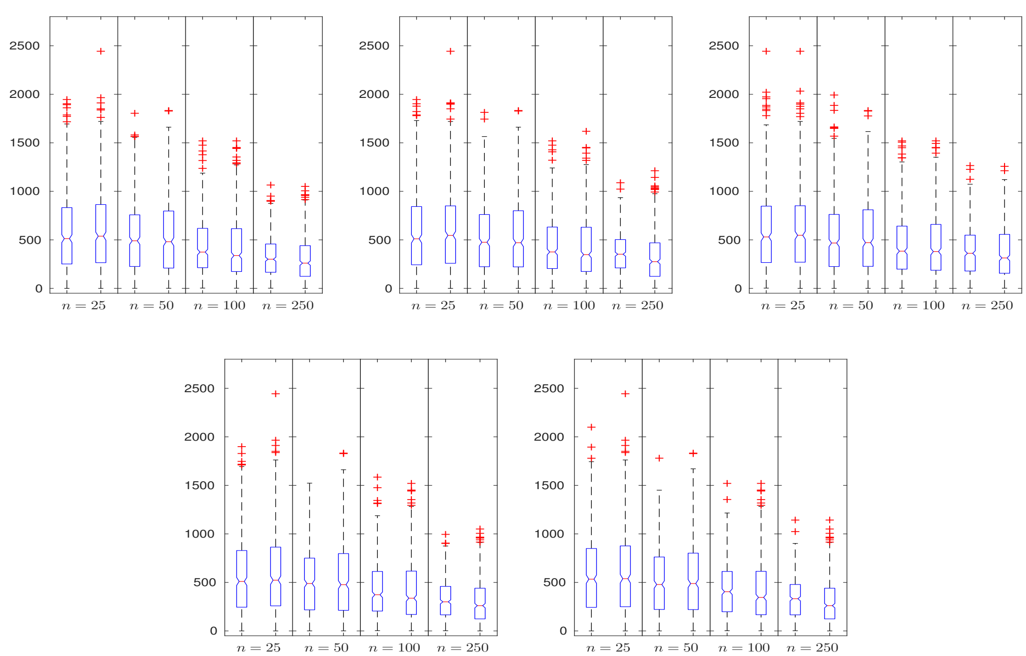

FIGURE 4.2. Boxplots comparing Δ^*_{wa} and Δ^*_{AIC} computed from the $\mathrm{CVaR}_{0.75}$-based optimisation cases. Each graph constitutes of four groups of boxplots that correspond to various sample sizes of n. The boxplot on the left/right-hand side represents $\Delta^*_{wa}/\Delta^*_{AIC}$. The top row boxplots are corresponding to distribution collections $\mathcal{M}_5$, $\mathcal{M}_4$ and $\mathcal{M}_2$, while the bottom row relates to $\mathcal{M}_4^*$ and $\mathcal{M}_2^*$, respectively.

is small. The last four rows displayed in Tables 4.7 and 4.8 summarise comparisons amongst

	$n=25$			$n=50$			$n=100$			$n=250$		
	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$
$S^*_{wc} \succ S^*_{wa}(\alpha=0.9)$	204	144	208	247	211	210	224	201	138	247	229	80
$S^*_{wa} \succ S^*_{wc}(\alpha=0.9)$	296	356	292	253	289	290	276	299	362	253	271	420
$S^*_{ad} \succ S^*_{wa}(\alpha=0.9)$	187	173	172	127	133	116	70	109	71	90	84	28
$S^*_{wa} \succ S^*_{ad}(\alpha=0.9)$	310	325	327	372	365	384	425	389	427	410	415	471
$S^*_{wa} \succ S^*_{AIC}(\alpha=0.9)$	114	142	153	71	108	156	24	88	165	3	92	200
$S^*_{AIC} \succ S^*_{wa}(\alpha=0.9)$	386	358	347	429	392	344	476	412	335	497	408	300
$S^*_{wa} \succ S^*_{AIC}(\alpha=0.2)$	235	235	243	267	286	284	229	254	254	210	223	251
$S^*_{AIC} \succ S^*_{wa}(\alpha=0.2)$	264	265	254	233	214	216	271	246	246	290	277	249

TABLE 4.7. Results when the PHT-based ($\alpha = 0.9$) Weighted Average Model is compared to the Worst-case, the Additive and the AIC models under various sample sizes n and collections of candidate models $\{\mathcal{M}_2, \mathcal{M}_4, \mathcal{M}_5\}$.

	$n=25$			$n=50$			$n=100$			$n=250$		
	$\mathcal{M}_5$	$\mathcal{M}^*_4$	$\mathcal{M}^*_2$	$\mathcal{M}_5$	$\mathcal{M}^*_4$	$\mathcal{M}^*_2$	$\mathcal{M}_5$	$\mathcal{M}^*_4$	$\mathcal{M}^*_2$	$\mathcal{M}_5$	$\mathcal{M}^*_4$	$\mathcal{M}^*_2$
$S^*_{wc} \succ S^*_{wa}(\alpha=0.9)$	204	263	218	247	284	250	224	288	298	247	309	337
$S^*_{wa} \succ S^*_{wc}(\alpha=0.9)$	296	237	282	253	216	250	276	212	202	253	191	163
$S^*_{ad} \succ S^*_{wa}(\alpha=0.9)$	187	145	219	127	114	270	70	70	346	90	130	470
$S^*_{wa} \succ S^*_{ad}(\alpha=0.9)$	310	352	279	372	382	227	425	425	154	410	370	30
$S^*_{wa} \succ S^*_{AIC}(\alpha=0.9)$	114	133	190	71	90	139	24	26	44	3	3	4
$S^*_{AIC} \succ S^*_{wa}(\alpha=0.9)$	386	367	310	429	410	361	476	474	456	497	497	496
$S^*_{wa} \succ S^*_{AIC}(\alpha=0.2)$	235	228	255	267	247	241	229	228	226	210	207	207
$S^*_{AIC} \succ S^*_{wa}(\alpha=0.2)$	264	272	245	233	253	258	271	272	274	290	293	293

TABLE 4.8. Results when the PHT-based Weighted Average Model is compared to the Worst-case, the Additive and the AIC models various sample sizes n and collection of candidate models $\{\mathcal{M}^*_2, \mathcal{M}^*_4, \mathcal{M}_5\}$.

optimal contracts found under PHT-based criterion with $\alpha = 0.9$ and 0.2. The performance of the Weighted Average Model (3.11) is rather weak when compared to the AIC Model when $\alpha = 0.9$. This outcome does not look surprising since $\frac{1}{\alpha}$ represents the risk aversion index, and

	$n=25$			$n=50$			$n=100$			$n=250$		
	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_4^*$
$S^*_{wwc} \succ S^*_{wa}$ $(l=2)$	172	179	196	202	212	192	149	203	149	191	221	203
$S^*_{wa} \succ S^*_{wwc}$ $(l=2)$	328	321	304	298	288	308	351	297	351	309	279	297
$S^*_{wwc} \succ S^*_{wa}$ $(l=3)$	182	191	148	182	219	139	128	197	101	150	186	168
$S^*_{wa} \succ S^*_{wwc}$ $(l=3)$	318	309	352	318	281	361	372	303	399	350	314	332
$S^*_{wwc} \succ S^*_{wa}$ $(l=4)$	161			174			95			132		
$S^*_{wa} \succ S^*_{wwc}$ $(l=4)$	339			326			405			368		

TABLE 4.9. Comparison between the PHT-based ($\alpha = 0.9$) solutions for various sample sizes n and collection of candidate models $\{\mathcal{M}_5, \mathcal{M}_4, \mathcal{M}_4^*\}$.

the greater this value is, the more risk aversion the decision-maker is. When α is close to one, the decision-maker acts less prudent, in which case robust optimal contracts are less of interest to the decision-maker. This is even further supported by our results when replicating the same experiment with a more risk-averse decision maker, i.e. α is reduced from 0.9 to 0.2, which could be seen in the last two rows of Tables 4.7 and 4.8. It is straightforward to notice that there is a significant improvement in the performance of our robust optimisation model, but unfortunately it is not sufficient enough to conclude that it outperforms the AIC Model.

Figure 4.3 illustrates the distributions of Δ^*_{wa} and Δ^*_{AIC} for the PHT-based case with $c = 0.2$. As before, the range of Δ^*_{wa} and Δ^*_{AIC} are very similar in most of the comparisons, especially when n is small, telling us that there is not enough evidence to say that the AIC Model provides a more robust solution than the Weighted Average Model (3.11).

The last set of results of the section considers the robustness of the SD-based optimal contracts where $b = 0.5$. The first eight rows in Tables 4.10 and 4.11 compare the Weighted Average Model (3.15) to the Weighted Worst-case Model (3.16) with $2 \le l \le m$, and the results are different than before. That is, the Weighted Worst-case Model is preferred for almost any sample size, but it is more clear when the sample size is small. In addition, the evidence tends to be more significant as l gets bigger, and hence, when our robust models are compared to the non-robust AIC model, we show only the comparison results relating to the Weighted Worst-case Model with $l = m$, which is displayed in the last two rows in Tables 4.10 and 4.11. Note that when $l = m$, the Weighted Worst-case Model is indeed the Additive Model (3.14). One may find that the AIC Model is only preferred over our robust Weighted Worst-case model when sample size is rather large, e.g. $n = 250$, otherwise the Weighted Worst-case model is

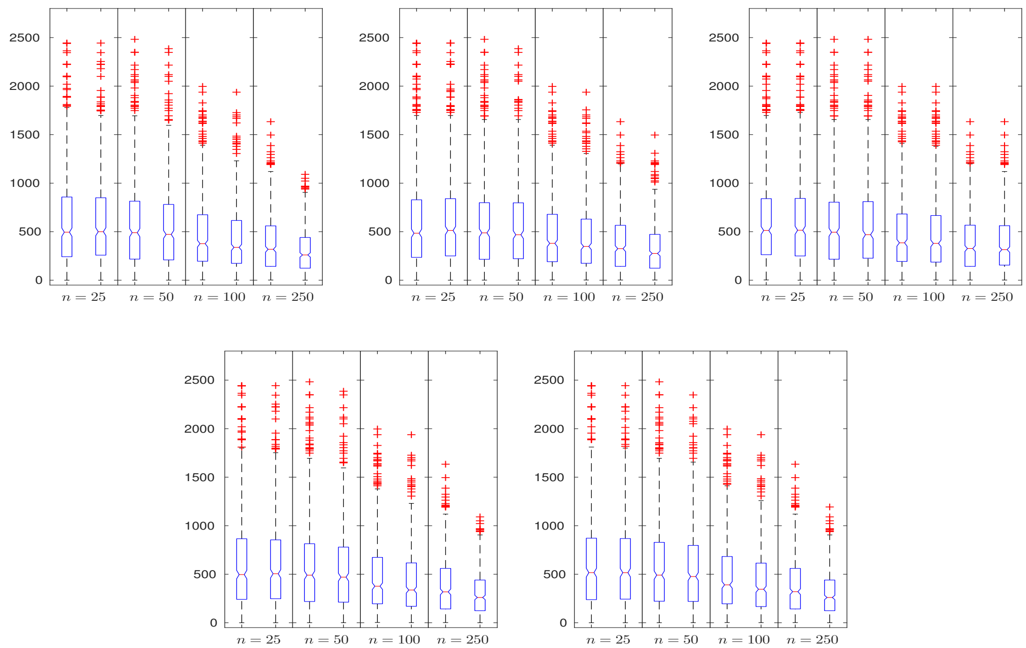

FIGURE 4.3. Boxplots comparing Δ^*_{wa} and Δ^*_{AIC} computed from the $PHT_{0.2}$-based optimisation cases. Each graph constitutes of four groups of boxplots that correspond to various sample sizes of n. The boxplot on the left/right-hand side represents $\Delta^*_{wa}/\Delta^*_{AIC}$. The top row boxplots are corresponding to distribution collections $\mathcal{M}_5$, $\mathcal{M}_4$ and $\mathcal{M}_2$, while the bottom row relates to $\mathcal{M}^*_4$ and $\mathcal{M}^*_2$, respectively.

recommended. This is consistent with the boxplot results displayed in Figure 4.4. These results could be explained by the risk measure choice, since standard deviation does not measure the tail of the distribution and therefore, the Weighted Worst-case Model overcomes this shortcoming. Once again, the sample size plays an important role and the AIC Model always leads to more robust solutions when data scarcity is not present.

It is also worth mentioning as a final remark that if we compare all the boxplots in Figures 4.1–4.4, the Δ^* resulted from the VaR-based optimisations tend to be smaller than those found under optimisations based on other risk measures, i.e. CVaR, PHT and SD, which could be explained by the robustness of VaR itself as a risk measure.

4.2. **Stability.** This section provides analyses on the stability of our empirical robust optimal insurance contracts. In order to avoid excessive repeats, we only report the stability of empirical solutions found from the most robust model as shown in Section 4.1, i.e. the Weighted Average Model for the VaR-, CVaR- and PHT-based cases and the Weighted Worst-case Model for the SD-based case. The scatter plots of $\mathbf{y}^*_{wa}$ and $\mathbf{y}^*_{wwc}$ against $\mathbf{x}$ are shown in Figure 4.5 for

	$n=25$			$n=50$			$n=100$			$n=250$		
	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_2$
$S^*_{wwc} \succ S^*_{wa}$ $(l=2)$	296	260	319	326	276	332	294	256	296	246	250	255
$S^*_{wa} \succ S^*_{wwc}$ $(l=2)$	204	240	181	174	224	168	206	244	204	254	250	245
$S^*_{wwc} \succ S^*_{wa}$ $(l=3)$	309	319		347	313		303	281		244	257	
$S^*_{wa} \succ S^*_{wwc}$ $(l=3)$	191	181		153	187		197	219		256	243	
$S^*_{wwc} \succ S^*_{wa}$ $(l=4)$	326	331		350	308		308	284		247	251	
$S^*_{wa} \succ S^*_{wwc}$ $(l=4)$	174	169		150	192		192	216		253	249	
$S^*_{wwc} \succ S^*_{wa}$ $(l=5)$	325			336			324			259		
$S^*_{wa} \succ S^*_{wwc}$ $(l=5)$	175			164			176			241		
$S^*_{ad} \succ S^*_{AIC}$	287	297	309	304	317	334	257	269	293	146	169	237
$S^*_{AIC} \succ S^*_{ad}$	213	203	191	196	183	166	243	231	207	354	331	263

TABLE 4.10. Comparison between the SD-based (b=0.5) solutions of (3.14), (3.15), (3.16) and the non-robust AIC model for various sample sizes n and collection of candidate models $\{\mathcal{M}_5, \mathcal{M}_4, \mathcal{M}_2\}$.

$n = 25, 100, 250$. It is observed that the VaR-based empirical solution mimics the functional form of $y^*_{wa,i} = c\Big(\big(x_i - d_1\big)_+ - \big(x_i - d_2\big)_+\Big)$, while the empirical solutions of all other cases mimic the functional form of $y^*_{wa,i} = c\big(x_i - d_1\big)_+$ and $y^*_{wwc,i} = c\big(x_i - d_1\big)_+$, where c, d_1 and d_2 are unknown parameters that can be estimated by *Ordinary Least Square (OLS)* regression fitting the functional forms to the corresponding data (x_i, y^*_i), $i = 1, 2, \ldots, n$. Recall that our numerical experiment contains 500 samples for each choice of sample size n. That is, there are 500 estimated pairs of the unknown parameters, $\left(\hat{c}, \hat{d}_1, \hat{d}_2\right)$, for each of $n = 25, 100, 250$, which is summarised in Table 4.12. Although variations exist in the mean values of $\left(\hat{c}, \hat{d}_1\right)$ for the CVaR-, PHT- and SD-based cases, it is noticed that the standard errors of $\left(\hat{c}, \hat{d}_1\right)$ has a decreasing trend as the sample size n grows. That is, we may conclude that the empirical solution of our Weighted Average Model for CVaR- and PHT-based cases and our Weighted Worst-case Model for SD-based case are stable and consistent. Unfortunately, such feature is not observed in the empirical solutions for the VaR-based cases. However, this should not become a major concern, as we have seen in Section 4.1 that our robust models are not the best options for solving VaR-based cases and AIC Model is recommended instead.

	$n=25$			$n=50$			$n=100$			$n=250$		
	$\mathcal{M}_5$	$\mathcal{M}_4^*$	$\mathcal{M}_2^*$	$\mathcal{M}_5$	$\mathcal{M}_4^*$	$\mathcal{M}_2^*$	$\mathcal{M}_5$	$\mathcal{M}_4^*$	$\mathcal{M}_2^*$	$\mathcal{M}_5$	$\mathcal{M}_4^*$	$\mathcal{M}_2^*$
$S^*_{wwc} \succ S^*_{wa}$ $(l=2)$	296	298	278	326	322	279	294	297	237	246	247	211
$S^*_{wa} \succ S^*_{wwc}$ $(l=2)$	204	202	222	174	178	263	206	203	263	254	253	289
$S^*_{wwc} \succ S^*_{wa}$ $(l=3)$	309	308		347	331		303	307		244	236	
$S^*_{wa} \succ S^*_{wwc}$ $(l=3)$	191	192		153	169		197	193		256	264	
$S^*_{wwc} \succ S^*_{wa}$ $(l=4)$	326	295		350	306		308	336		247	271	
$S^*_{wa} \succ S^*_{wwc}$ $(l=4)$	174	205		150	194		192	164		253	229	
$S^*_{wwc} \succ S^*_{wa}$ $(l=5)$	325			336			324			259		
$S^*_{wa} \succ S^*_{wwc}$ $(l=5)$	175			164			176			241		
$S^*_{ad} \succ S^*_{AIC}$	287	293	298	304	301	301	257	267	236	146	170	186
$S^*_{AIC} \succ S^*_{ad}$	213	207	202	196	199	199	243	233	264	354	330	314

TABLE 4.11. Comparison between the SD-based (b=0.5) solutions of (3.14), (3.15), (3.16) and the non-robust AIC model for various sample sizes n and collection of candidate models $\{\mathcal{M}_5, \mathcal{M}_4^*, \mathcal{M}_2\}$.

		CVaR Case		PHT Case		SD Case	
		$\hat{c}$	$\hat{d}_1$	$\hat{c}$	$\hat{d}_1$	$\hat{c}$	$\hat{d}_1$
$n=25$	Mean	0.916	2892.3	1.000	4460.0	0.9287	3343.0
	(Standard Error)	(0.1650)	(1908.6)	(0.7195)	(5268.6)	(0.1011)	(2481.6)
$n=100$	Mean	0.8940	3147.0	1.000	4531.9	0.9075	3433.8
	(Standard Error)	(0.1705)	(1222.7)	(0.0000)	(2787.0)	(0.0781)	(1412.0)
$n=250$	Mean	0.8888	3212.3	1.000	4520.2	0.8925	3359.1
	(Standard Error)	(0.1410)	(795.44)	(0.0000)	(1806.9)	(0.06210)	(933.07)

TABLE 4.12. Summary of mean and standard errors of $\left(\hat{c}, \hat{d}_1\right)$ for CVaR-, PHT- and SD-based cases with various sample size n.

5. CONCLUSIONS

Robust optimal insurance contracts have been investigated by carrying out many numerical experiments under various risk-based decisions. It is concluded that the sample size plays a major role in the sense that, whenever data scarcity is not present, the AIC Model is preferred and there is a need to focus on available statistical methods in order to find the most robust optimal

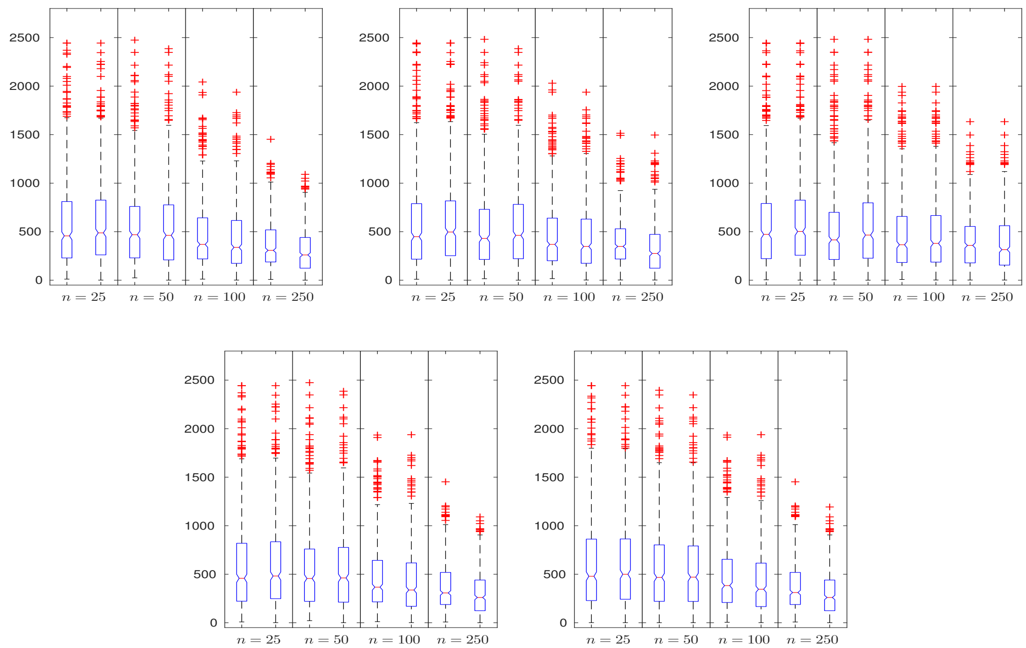

FIGURE 4.4. Boxplots comparing Δ^*_{wa} and Δ^*_{AIC} computed from the SD-based ($b = 0.5$) optimisation cases. Each graph constitutes of four groups of boxplots that correspond to various sample sizes of n. The boxplot on the left/right-hand side represents $\Delta^*_{wa}/\Delta^*_{AIC}$. The top row boxplots are corresponding to distribution collections $\mathcal{M}_5$, $\mathcal{M}_4$ and $\mathcal{M}_2$, while the bottom row relates to $\mathcal{M}^*_4$ and $\mathcal{M}^*_2$, respectively.

decision. If small samples are available, then either the Weighted Average Model or Weighted Worst-case Model should be considered instead of trying to identify the "best" statistical tool to estimate the unknown risk model. Our numerical experiments have shown that whenever the decision-maker has a particular interest in the tail distribution, i.e. the decisions are based on VaR, CVaR or PHT, the Weighted Average Model produces the most robust solutions whenever the available sample is relatively small. On the other hand, the Weighted Worst-case Model leads to the most robust optimal solution if the decision-maker has little interest in the tail risk and thus, such risk preferences require a robust method that puts more weight on the worst cases. These conclusions reiterate once again that one should be very careful when robust optimal decisions are sought and one should first understand the features of the objective function and the size of the available data, and then decide whether robust optimisation or statistical inferences are the way forward.

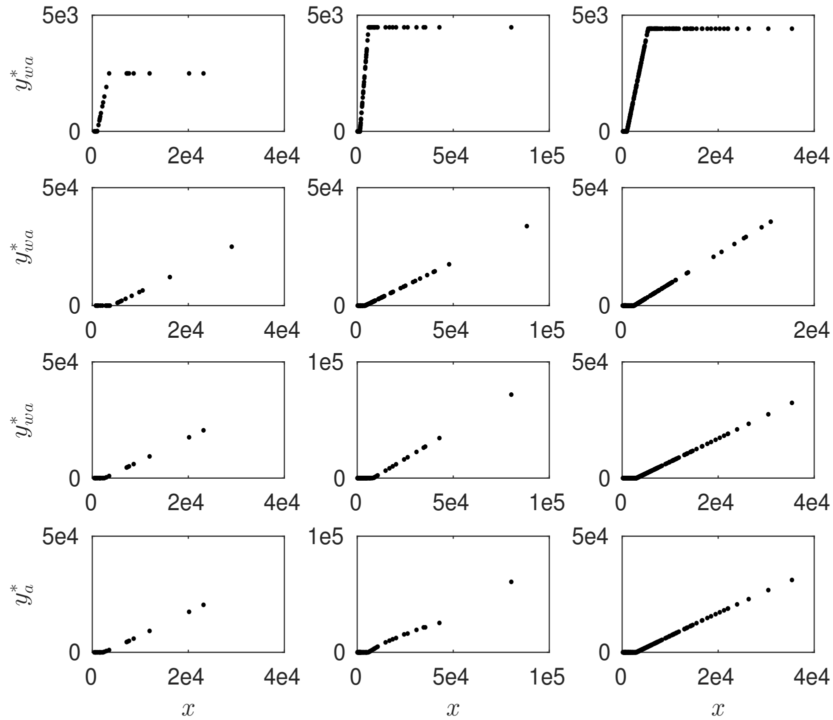

FIGURE 4.5. Scatter plots of empirical robust optimal insurance contracts found from various robust optimisation models and sample sizes. The plots in each row (from top to bottom) correspond to the VaR-, CVaR- and PHT-based Weighted Average Models and the SD-based Additive Model, respectively. The plots in each column (from left to right) correspond to the sample size of $n = 25$, 100 and 250, respectively.

6. PROOFS

Proof of Proposition 2.1. The reformulation (2.6) tells us that minimising (2.5) over $\mathcal{A}$ can be written as follows

$$\min_{(\mathbf{t},s)\in\mathcal{A}\times\Re}\left\{s+\frac{1}{l}\sum_{i=1}^{m}\Big(f\big(\mathbf{t};\boldsymbol{\omega}_i\big)-s\Big)_+\right\}, \tag{6.1}$$

and we show that solving the above problem is equivalent to solving the optimisation problem (2.7). Let us denote the optimal solution to (2.7) as $(\mathbf{t}^*, s^*, \mathbf{u}^*)$. It is noticed that the objective function in (2.7) is increasing in u_i for all $i \in \mathcal{M}$, and therefore, constraints $f\big(\mathbf{t};\boldsymbol{\omega}_i\big) \le s+u_i$ and $\mathbf{0} \le \mathbf{u}$ ensure that $u_i^* = \Big(f\big(\mathbf{t}^*;\boldsymbol{\omega}_i\big) - s^*\Big)_+$ for all $i \in \mathcal{M}$. Consequently, $(\mathbf{t}^*, s^*)$ is also feasible to the problem (6.1). Suppose that $(\mathbf{t}^*, s^*)$ is not the optimal solution to (6.1), then there must exist another feasible solution $\big(\mathbf{t}', s'\big)$ such that

$$s' + \frac{1}{l}\sum_{i=1}^{m}\Big(f\big(\mathbf{t}';\boldsymbol{\omega}_i\big) - s'\Big)_+ < s^* + \frac{1}{l}\sum_{i=1}^{m}\Big(f\big(\mathbf{t}^*;\boldsymbol{\omega}_i\big) - s^*\Big)_+ = s^* + \frac{1}{l}\sum_{i=1}^{m} u_i^*. \tag{6.2}$$

Note that $(\mathbf{t}', s', \mathbf{u}')$ with $u_i' = \Big(f\big(\mathbf{t}';\boldsymbol{\omega}_i\big) - s'\Big)_+$ for all $i \in \mathcal{M}$ is also feasible to (2.7). However,

$$s' + \frac{1}{l}\sum_{i=1}^{m} u_i' < s^* + \frac{1}{l}\sum_{i=1}^{m} u_i^*$$

is implied by (6.2), which contradicts the assumption that $(\mathbf{t}^*, s^*, \mathbf{u}^*)$ is the optimal solution to the optimisation problem (2.7). As a result, the optimal solution to (2.7) must also solve the problem (6.1).

On the other hand, suppose that $(\mathbf{t}^*, s^*)$ is the optimal solution to (6.1). Then, $(\mathbf{t}^*, s^*, \mathbf{u}^*)$ with $u_i^* = \Big(f\big(\mathbf{t}^*;\boldsymbol{\omega}_i\big) - s^*\Big)_+$ for all $i \in \mathcal{M}$ is also feasible to (2.7). If $(\mathbf{t}^*, s^*, \mathbf{u}^*)$ is not an optimal solution to (2.7), there must exist another feasible solution $(\mathbf{t}', s', \mathbf{u}')$ such that

$$s' + \frac{1}{l}\sum_{i=1}^{m} u_i' < s^* + \frac{1}{l}\sum_{i=1}^{m} u_i^* = s^* + \frac{1}{l}\sum_{i=1}^{m}\Big(f\big(\mathbf{t}^*;\boldsymbol{\omega}_i\big) - s^*\Big)_+. \tag{6.3}$$

Since the constraints $f\big(\mathbf{t};\boldsymbol{\omega}_i\big) \le s + u_i$ and $\mathbf{0} \le \mathbf{u}$ in (2.7) will ensure $u_i' = \Big(f\big(\mathbf{t}';\boldsymbol{\omega}_i\big) - s'\Big)_+$, $(\mathbf{t}', s')$ is also feasible to (6.1) with

$$s' + \frac{1}{l}\sum_{i=1}^{m} u_i' < s^* + \frac{1}{l}\sum_{i=1}^{m} u_i^*$$

implied by (6.3), which then contradicts the assumption of $(\mathbf{t}^*, s^*)$ being the optimal solution to (6.1). That is, the optimal solution to (6.1) must also solve the optimisation problem (2.7). The proof is completed by combining both arguments. □

Proof of Theorem 3.1. Let us first show that an optimal solution $\mathbf{x}^*$ of (3.17) must be Pareto optimal when the optimal objective function value in (3.18) is zero. If $\mathbf{x}^*$ is not Pareto optimal, then there must exist another feasible solution $\hat{\mathbf{y}}$ of (3.17) such that $f_k\big(\hat{\mathbf{y}}\big) \le f_k\big(\mathbf{x}^*\big)$ for all $k \in \mathcal{M}$ with at least one inequality sign being strict. Thus, $\hat{\mathbf{y}}$ is feasible in (3.18) and

$$\sum_{k\in\mathcal{M}}\Big(f_k\big(\hat{\mathbf{y}}\big) - f_k\big(\mathbf{x}^*\big)\Big) < 0,$$

which contradicts the statement that the optimal objective function value of (3.18) is zero. Thus, $\mathbf{x}^*$ must be Pareto optimal.

Next, we show that when the optimal objective function value of (3.18) is negative, any optimal solution $\mathbf{y}^*$ of (3.18) solves (3.17) as well and is Pareto optimal. Now, $f_k\big(\mathbf{y}^*\big) \le f_k\big(\mathbf{x}^*\big)$ for any $k \in \mathcal{M}$, since $\mathbf{y}^*$ is feasible in (3.18), which in turn gives that $f^{(k)}\big(\mathbf{y}^*\big) \le f^{(k)}\big(\mathbf{x}^*\big)$ for any $k \in \mathcal{M}$. The latter and the fact λ_k's are positive imply that $\mathbf{y}^*$ must solve (3.17), since $\mathbf{x}^*$

solves (3.17). Assume now that $\mathbf{y}^*$ is not Pareto optimal. Therefore, there must exist another feasible solution $\hat{\mathbf{y}}$ of (3.17) such that $f_k(\hat{\mathbf{y}}) \leq f_k(\mathbf{y}^*)$ for all $k \in \mathcal{M}$ with at least one inequality sign being strict. Consequently, $\hat{\mathbf{y}}$ is feasible in (3.18) and

$$\sum_{k\in\mathcal{M}} \Big(f_k(\hat{\mathbf{y}}) - f_k(\mathbf{x}^*)\Big) < \sum_{k\in\mathcal{M}} \Big(f_k(\mathbf{y}^*) - f_k(\mathbf{x}^*)\Big),$$

which contradicts the fact that $\mathbf{y}^*$ is an optimal solution of (3.18). Therefore, $\mathbf{y}^*$ must be Pareto optimal. The proof is now complete. □

References

Alizadeh, F. and Goldfarb, D. (2003). "Second-order cone programming." *Mathematical Programming*, 95(1), 3–51.

Arrow, K. J. (1963). "Uncertainty and the welfare economics of medical care", *American Economic Review*, 53(3), 941–973.

Asimit, A. V., Badescu, A. M. and Verdonck, T. (2013). "Optimal risk transfer under quantile-based risk measures", *Insurance: Mathematics and Economics*, 53, 252–265.

Asimit, V.A., Chi, Y. and Hu, J. (2015). "Optimal non-life reinsurance under Solvency II regime", *Insurance: Mathematics and Economics*, 65, 227–237.

Asimit, V., Bignozzi, V., Cheung, K.C., Hu, J. and Kim, E.-S. (2017). "Robust and Pareto optimality of insurance contracts", *European Journal of Operational Research*, 262(2), 720–732.

Asimit, A.V. and Boonen, T.J. (2018). "Insurance with multiple insurers: a game-theoretic approach", *European Journal of Operational Research*, in press, DOI: 10.1016/j.ejor.2017.12.026.

Asimit, A.V., Gao, T., Hu, J. and Kim, E.-S. (2018). "Optimal risk transfer: a numerical optimisation approach", *North American Actuarial Journal*, 22(3), 341–364.

Balbás, A., Balbás, B. and Heras, A. (2009). "Optimal reinsurance with general risk measures", *Insurance: Mathematics and Economics*, 44, 374–384.

Balbás, A., Balbás, B. and Heras, A. (2011). "Stable solutions for optimal reinsurance problems involving risk measures", *European Journal of Operational Research*, 214, 796–804.

Balbás, A., Balbás, B., Balbás, R. and Heras, A. (2015). "Optimal reinsurance under risk and uncertainty", *Insurance: Mathematics and Economics*, 60, 61–74.

Ben-Tal, A., El Ghaoui, L. and Nemirovski, A. (2009). "Robust optimization", *Princeton University Press*, New Jersey.

Ben-Tal, A. and Nemirovski, A. (2002). "Robust optimization – methodology and applications", *Mathematical Programming*, 92(3), 453–480.

Ben-Tal, A. and Nemirovski, A. (2008). "Selected topics in robust convex optimization", *Mathematical Programming*, 112(1), 125–158.

Bertsimas, D., Brown, D.B. and 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*, 597–610.

Cai, J., Tan, K.S., Weng, C. and Zhang, Y. (2008)."Optimal reinsurance under VaR and CTE risk measures", *Insurance: Mathematics and Economics*, 43(1), 185–196.

Cai, J. and Weng, C.G. (2016). "Optimal reinsurance with expectile", *Scandinavian Actuarial Journal*, 7, 624–645.

Chassein, A. and Goerigk, M. (2016). "Performance analysis in robust optimization", *Robustness Analysis in Decision Aiding, Optimization and Analytics*, International Series in Operations Research & Management Science, Book Series, vol 241, 145–170.

Cheung, K.C., Sung, K.C.J., Yam, S.C.P. and Yung, S.P. (2014). "Optimal reinsurance under general law-invariant risk measures", *Scandinavian Actuarial Journal*, 1, 72–91.

Chi, Y. and Tan, K.S. (2011)."Optimal reinsurance under VaR and CVaR risk measures: a simplified approach", *Astin Bulletin*, 41(2), 487–509.

European Commission (2010). "QIS5 Technical Specifications", available at: `http://eiopa.europa.eu/Publications/QIS/QIS5-technical_specifications_20100706.pdf`.

Gabrel, V., Murat, C. and Thiele, A. (2014). "Recent advances in robust optimization: an overview", *European Journal of Operational Research*, 235(3), 471–483.

Ghossoub, M. (2019). "Budget-constrained optimal insurance without the nonnegativity constraint on indemnities", *Insurance: Mathematics and Economics*, 84, 22–39.

Gneiting, T. (2011). "Making and evaluating point forecasts", *Journal of the American Statistical Association*, 106(494), 746–762.

Goovaerts, M., Kaas, R. and Laeven, R. (2011). "Worst case risk measurement: Back to the future?", *Insurance: Mathematics and Economics*, 49, 380–392.

Huang, D., Zhu, S., Fabozzi, F.J. and Fukushima, M. (2010). "Portfolio selection under distributional uncertainty: a relative robust CVaR approach", *European Journal of Operational Research*, 203, 185–194.

Lamber, N. S., Pennock, D. M. and Shoham, Y. (2008). "Eliciting properties of probability distributions", *Proceedings of the 9th ACM Conference on Electronic Commerce*, 129–138.

Lu, Z., Liu, L., Shen, Q. and Meng, L. (2014). "optimal reinsurance under VaR and CTE risk measures when ceded loss function is concave", *Communications in Statistics - Theory and Methods*, 43(15), 3223–3247.

Lobo, M.S., Vandenberghe, L., Boyd, S. and Lebret, H. (1998). "Applications of second-order cone programming", *Linear Algebra and its Applications*, 284(1998), 193–228.

Rockafeller, R.T. and Uryasev, S. (2000). "Optimisation of Conditional Value-at-Risk", *Journal of Risk*, 2, 21–41.

Satchell, S. (2010). "Optimizing optimization: The next generation of optimization applications and theory", *Academic Press*, London.

Tan, K. and Weng, C., (2014). "Empirical approach for optimal reinsurance design", *North American Actuarial Journal*, 18(2), 315–342.

Wang, S., Young, V. R. and Panjer, H. H. (1997). "Axiomatic characterization of insurance prices", *Insurance: Mathematics and Economics*, 21(2), 173–183.

Zhu, S. and Fukushima, M. (2009). "Worst-case Conditional Value-at-Risk with application to robust portfolio management", *Operations Research*, 57(5), 1155–1168.

Zymler, S., Kuhn, D. and Rustem, B. (2013). "Worst-case Value-at-Risk of nonlinear portfolios", *Management Science*, 59(1), 172–188.