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
Cass Business School, City, University of London, London, EC1Y 8TZ, United Kingdom. E-mail: Junlei.Hu.1@city.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

January 22, 2018

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 could be efficiently solved. The 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

\(^1\)Corresponding author. Phone: +861064494702. Fax: +861064495024.
objective function and the size of 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, but the same problem is known as the insurance demand problem in insurance economics field. 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 more simple, 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, standard robust methodologies are also available in the absence of data scarcity. 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. The performance of our robust optimisation models are empirically evaluated via solving Second Order Conic Programming (SOCP) instances, which could be efficiently solved; for more detailed examples of how to reduce complex optimal insurance problems to SOCP instances see Tan and Weng (2014) and Asimit et al. (2018), while Pareto optimality problem examples appear in Asimit and Boonen (2018).

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 
\( f(\cdot; \omega) : \mathcal{A} \to \mathbb{R} \) 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:

\[
\min_{\mathcal{A}} \sup_{\omega \in \mathcal{U}} f(\cdot; \omega),
\]

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 decisions. Most of the existing papers in the field focus on parameter risk and the parameter uncertainty set is defined as an uncertainty box, i.e. the parameters belong to a hypercube, ellipsoidal or polyhedral uncertainty set etc (amongst others, see Ben-Tal and Nemirovski, 2002 and 2008, Ben-Tal et al., 2009, Zhu and Fukushima, 2009 and Zymler et al., 2013). While one could argue in infinite ways which uncertainty box is more appropriate, it is an irrefutable fact that none of these are useful to incorporate the model error, i.e. when \( \mathcal{U} \) is a finite collection of parameteric and/or non-parametric models.

This approach is well-known in finance and insurance applications, where stochastic models are constructed as potential candidates to represent the unknown “true” model; for example, see Zhu and Fukushima (2009), Huang et al. (2010) and Asimit et al. (2017 and 2018), where one should note that the first two papers considered a convex hull of the candidate models. Note that this approach produces a large uncertainty set that may be affect the robust optimal decision and therefore, it would be better to initiate a non-convex uncertainty set that is only composed of the possible models as explained in Asimit et al. (2017). Specifically, if \( \mathcal{U} = \{\omega_k, k \in \mathcal{M}\} \), where \( \mathcal{M} := \{1, 2, \ldots, m\} \), then (2.1) becomes

\[
\min_{\mathcal{A}} \max_{k \in \mathcal{M}} f(\cdot; \omega_k).
\]

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

\[
\min_{\mathcal{A}} \max_{k \in \mathcal{M}} f(\cdot; \omega_k) - f_k^*, \quad \text{where} \quad f_k^* = \min_{t \in \mathcal{A}} f(\cdot; \omega_k) \quad \text{for all } k \in \mathcal{M}.
\]

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 $\lambda_k$,

$$
\min_{t \in A} \sum_{k \in M} \lambda_k f(t; \omega_k),
$$

(2.4)

where $\lambda \geq 0$ and $1^T \lambda = 1$. 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 scenario type

$$
f(t; \omega_1, \ldots, \omega_m, l) := \frac{1}{l} \sum_{i=1}^l f^{(i)}(t; \omega_1, \ldots, \omega_m, l), \ l \in M,
$$

(2.5)

where $f^{(i)}(; \omega_1, \ldots, \omega_m, l)$ is the $i^{th}$ upper order statistics of $\{f(\cdot; \omega_k), k \in M\}$, i.e.

$$f^{(i)}(t; \omega_1, \ldots, \omega_m, l) = f(t; \omega_{\sigma(i)}) \text{ such that } f(t; \omega_{\sigma(1)}) \geq \ldots \geq f(t; \omega_{\sigma(l)})$$

with $\sigma$ being a permutation of $M$. 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 $t \in A$, (2.5) could be reformulated in the following fashion

$$f(t; \omega_1, \ldots, \omega_m, l) = \min_{s \in \mathbb{R}} \left\{ s + \frac{1}{l} \sum_{i \in M} (f(t; \omega_i) - s)_+ \right\}, \text{ where } (t)_+ = \max(t, 0).$$

(2.6)

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

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

where $a^{(i)}$ represents the $i^{th}$ upper order statistics of $\{a_k, k \in M\}$. 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 $s \in (a_{j+1}, a_j)$, where $0 \leq j \leq m$, the following are true:

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

ii) If $j < l$, then we have that

$$s + \frac{1}{l} \sum_{i \in M} (a_i - s)_+ = \frac{1}{l} \left( \sum_{i=1}^j a_i + (l - j)s \right) \geq \frac{1}{l} \left( \sum_{i=1}^j a_i + (l - j)a_{j+1} \right) \geq \frac{1}{l} \sum_{i=1}^l a_i.$$
iii) If \( j > l \), then we have that
\[
s + \frac{1}{l} \sum_{i \in M} (a_i - s) = \frac{1}{l} \left( \sum_{i=1}^{j} a_i - (j - l)s \right) \geq \frac{1}{l} \left( \sum_{i=1}^{j} a_i - (j - l)a_j \right) = \sum_{i=1}^{l} a_i.
\]
As a result, \( s^* \in (a_{l+1}, a_l] \), which in turn justifies our claim.

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

**Proposition 2.1.** Optimising (2.6) over a convex set \( t \in A \), i.e.
\[
\min_{(t,s) \in A \times \mathbb{R}} \left\{ s + \frac{1}{l} \sum_{i \in M} (f(t; \omega_i) - s) \right\},
\]
is equivalent to solving
\[
\min_{(t,s,u) \in A \times \mathbb{R} \times \mathbb{R}^m} \left\{ s + \frac{1}{l} T u \right\}, \text{ s.t. } 0 \leq u_i, f(t; \omega_i) \leq s + u_i, \forall i \in M. \quad (2.7)
\]
The computational advantage of (2.7) is conspicuous, since most of the terms are linear. Specifically, if \( f(\cdot; \omega_i) \) are SOCP representable for all \( i \in M \), 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, \( I, R \in C^{co} \) where
\[
C^{co} = \{ f \text{ is non-decreasing } | 0 \leq f(x) \leq x, |f(x) - f(y)| \leq |x - y| \text{ for all } x, y \in \mathbb{R} \}.
\]

Assume that any feasible reinsurance contract satisfies \( \Phi(R[X]; \mathcal{P}) \leq P \leq \bar{P} \), 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, \( \Phi(R[X]; \mathcal{P}) \leq P \), 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 \( \Phi(R[X] - P; \mathcal{P}) \leq 0 \). The latter is equivalent to \( \Phi(R[X]; \mathcal{P}) \leq P \), if \( \Phi(0; \mathcal{P}) = 0 \) and \( \Phi \) 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 \( \Omega \). Then, for any \( X \in \mathcal{X} \) and \( a \in \mathbb{R} \), \( \Phi : \mathcal{X} \to \mathbb{R} \) is a translation invariant risk measure if \( \Phi(X + a) = \Phi(X) + a \).
Now, when the probability measure \( P \) 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 \( \{P_1, P_2, \ldots, P_m\} \). Then, the feasibility constraint becomes \( \Phi(R[X]; P_k) \leq P \leq \bar{P} \) 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 \( \rho \) and thus, its objective under the \( k^{th} \) model is \( \rho(X - R[X] + P; P_k) \), which reduces to \( \rho(X - R[X]; P_k) + P \) if \( \rho \) 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) \textit{wc-type} as defined in (2.2)

\[
\min_{(R, P) \in C_{\alpha} \times \mathbb{R}} \left\{ \max_{k \in \mathcal{M}} \rho(X - R[X]; P_k) + P \right\} \quad \text{s.t.} \quad \Phi(R[X]; P_k) \leq P \leq \bar{P} \quad \forall k \in \mathcal{M}. \tag{2.8}
\]

B) \textit{ad-type} as given in (2.4)

\[
\min_{(R, P) \in C_{\alpha} \times \mathbb{R}} \left\{ \sum_{k \in \mathcal{M}} \rho(X - R[X]; P_k) + P \right\} \quad \text{s.t.} \quad \Phi(R[X]; P_k) \leq P \leq \bar{P} \quad \forall k \in \mathcal{M}. \tag{2.9}
\]

C) \textit{wa-type} as defined in (2.4)

\[
\min_{(R, P) \in C_{\alpha} \times \mathbb{R}} \left\{ \sum_{k \in \mathcal{M}} \lambda_k \rho(X - R[X]; P_k) + P \right\} \quad \text{s.t.} \quad \Phi(R[X]; P_k) \leq P \leq \bar{P} \quad \forall k \in \mathcal{M}. \tag{2.10}
\]

D) \textit{wwc-type} as given in (2.5)

\[
\min_{(R, P) \in C_{\alpha} \times \mathbb{R}} \left\{ \frac{1}{l} \sum_{i=1}^{l} \rho(X - R[X]; P_{\sigma(i)}) + P \right\} \quad \text{s.t.} \quad \Phi(R[X]; P_k) \leq P \leq \bar{P} \quad \forall k \in \mathcal{M}, \tag{2.11}
\]

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

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

3. Empirical Formulations

3.1. \textbf{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 \( x := (x_1, x_2, \ldots, x_n)^T \). 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 \( y := (y_1, y_2, \ldots, y_n)^T \) 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, \( p_k := (p_{1k}, p_{2k}, \ldots, p_{nk})^T \), 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 
\textit{Value-at-Risk (VaR)} and \textit{Conditional Value-at-Risk (CVaR)}. The VaR of a generic loss variable \( Z > 0 \) with confidence level \( \alpha \in (0, 1) \) is defined as
\[
\text{VaR}_\alpha(Z; \mathcal{P}) := \inf_{y \in \mathbb{R}} \left\{ \mathcal{P}(Z \leq z) \geq \alpha \right\},
\]
while the CVaR is given as (see Rockafeller and Uryasev, 2000):
\[
\text{CVaR}_\alpha(Z; \mathcal{P}) := \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{1-\alpha} \mathbb{E}_\mathcal{P}(Z - t)_+ \right\}.
\]
By definition, \( \mathbb{E}_\mathcal{P}(\cdot) \) represents the expectation with respect to \( \mathcal{P} \).

Recall that \( R \in \mathcal{C}^{co} \), which implies that \( x, y \) and \( x - y \) are all non-decreasingly ordered. Therefore the empirical measure of \( \text{VaR}_\alpha(X - R[X]; \mathcal{P}_k) \) 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 \( \text{CVaR}_\alpha(X - R[X]; \mathcal{P}_k) \) becomes \( \phi_k^T x - \phi_k^T y \), where \( \phi_k := (\phi_{1k}, \phi_{2k}, \ldots, \phi_{nk})^T \) with
\[
\phi_{ik} = g \left( 1 - \sum_{j=1}^{i-1} p_{jk} \right) - g \left( 1 - \sum_{j=1}^i p_{jk} \right), \quad 1 \leq i \leq n, \ k \in \mathcal{M}
\]
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 already mentioned in Section 2.2 that \( \rho \) and \( \Phi \) are assumed to be translation invariant risk measures in this paper. Without loss of generality, the expected value premium principle is assumed, i.e. \( \Phi(\cdot; \mathcal{P}) = (1 + \theta)\mathbb{E}_\mathcal{P}(\cdot) \) with \( \theta > 0 \). In turn, the premium constraints become:
\[
(1 + \theta)p_k^T y \leq P \leq \mathcal{P}, \quad \forall k \in \mathcal{M}.
\]
It is important to mention that our numerical experiments had shown that the choice of premium principle does not have an impact on our conclusions and for this reason, the numerical analysis from this paper is focused on examples with the expected value premium principle being in force. Recall that \( X > 0 \) and \( I, R \in \mathcal{C}^{co} \), which is equivalent to
\[
0 \leq y \leq x, \quad 0 \leq Ay \leq Ax,
\]
where $A$ is an $n$-by-$n$ matrix given by

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

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. $\rho(\cdot; \mathcal{P}) = \text{VaR}_\alpha(\cdot; \mathcal{P})$. Since $X - R[X] \in \mathcal{C}^{co}$, we have that $\text{VaR}_\alpha(X - R[X]; \mathcal{P}_k) = x_{p(k)} - y_{p(k)}$ for all $k \in \mathcal{M}$. Therefore,

A) The \textit{wc-type} optimisation problem from (2.8) becomes

$$\min_{(y, P, r) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}} \ r \quad \text{s.t.} \quad x_{p(k)} - y_{p(k)} + P \leq r, \ \forall \ k \in \mathcal{M}, \ (3.2) \text{ and } (3.3) \text{ hold.} \quad (3.4)$$

B) The \textit{ad-type} optimisation problem from (2.9) becomes

$$\min_{(y, P) \in \mathbb{R}^n \times \mathbb{R}} \ \left\{ \sum_{k \in \mathcal{M}} (x_{p(k)} - y_{p(k)}) + P \right\} \quad \text{s.t.} \ (3.2) \text{ and } (3.3) \text{ hold.} \quad (3.5)$$

C) The \textit{wa-type} optimisation problem from (2.10) becomes

$$\min_{(y, P) \in \mathbb{R}^n \times \mathbb{R}} \ \left\{ \sum_{k \in \mathcal{M}} \lambda_k (x_{p(k)} - y_{p(k)}) + P \right\} \quad \text{s.t.} \ (3.2) \text{ and } (3.3) \text{ hold.} \quad (3.6)$$

D) The \textit{wwc-type} optimisation problem from (2.11) becomes

$$\min_{(y, P, r, s, u) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^m} \ \{ r + P \} \quad \text{s.t.} \quad s + \frac{1}{l} 1^T u \leq r, \ 0 \leq u, \ (3.2) \text{ and } (3.3) \text{ hold,}$$

$$x_{p(k)} - y_{p(k)} - s \leq u_k, \ \forall \ k \in \mathcal{M}. \quad (3.7)$$

The epigraph form from (3.4) is a standard reformulation in optimisation, while (3.5) and (3.6) 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. $\rho(\cdot; \mathcal{P}) = \text{CVaR}_\alpha(\cdot; \mathcal{P})$. Since $X - R[X] \in \mathcal{C}^{co}$, we have that $\text{CVaR}_\alpha(X - R[X]; \mathcal{P}_k) = \phi_k^T x - \phi_k^T y$, $\forall \ k \in \mathcal{M}$, by keeping in mind (3.1). Therefore, (2.8)–(2.11) are equivalent to solving the following optimisation problems:

A) The \textit{wc-type} optimisation problem from (2.8) becomes

$$\min_{(y, P, r) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}} \ r \quad \text{s.t.} \quad \phi_k^T x - \phi_k^T y + P \leq r, \ \forall \ k \in \mathcal{M}, \ (3.2) \text{ and } (3.3) \text{ hold.} \quad (3.8)$$
B) The *ad-type* optimisation problem from (2.9) becomes

$$\min_{(y,P) \in \mathbb{R}^n \times \mathbb{R}} \left\{ \sum_{k \in M} \left( \phi_k^T x - \phi_k^T y \right) + P \right\} \quad \text{s.t. (3.2) and (3.3) hold.} \tag{3.9}$$

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

$$\min_{(y,P) \in \mathbb{R}^n \times \mathbb{R}} \left\{ \sum_{k \in M} \lambda_k \left( \phi_k^T x - \phi_k^T y \right) + P \right\} \quad \text{s.t. (3.2) and (3.3) hold.} \tag{3.10}$$

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

$$\min_{(y,P,r,s,u) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^m} \{ r + P \} \quad \text{s.t.} \quad s + \frac{1}{l} I^T u \leq r, \ 0 \leq u, \ (3.2) \text{ and } (3.3) \text{ hold,}$$

$$\phi_k^T x - \phi_k^T y - s \leq u_k, \ \forall \ k \in M.$$ \tag{3.11}

The epigraph form from (3.8) is a standard reformulation in optimisation, while (3.9) and (3.10) are straightforward reformulations that do not require any additional work.

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

$$\text{PH}T^\alpha(X - R[X]; \mathcal{P}_k) = \phi_k^T x - \phi_k^T y, \quad \forall \ k \in M,$$

where $\phi_k$ are defined as in (3.1) with $g(t) = t^\alpha$. Therefore, the robust optimisation problems (2.8)–(2.11) are precisely as in (3.8)–(3.11), but with different parameters $\phi_k$'s.

The final case is when the insurance buyer orders its risk preferences as via the standard deviation SD risk measure, i.e. $\rho(\cdot; \mathcal{P}) = E_{\mathcal{P}}(\cdot) + b Sd(\cdot; \mathcal{P})$ 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 $Sd(Z; \mathcal{P}) = \|Qz\|$, where $Q$ is a $n \times n$ matrix with its $(j_1, j_2)$-th element to be $q_{j_1,j_2} = \sqrt{p_{j_1}}(1_{j_1 = j_2} - p_{j_2})$ for all $1 \leq j_1, j_2 \leq n$. 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

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

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

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

$$\min_{(y,P,r) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}} \ r \quad \text{s.t.} \quad p_k^T(x - y) + b\|Q_k(x - y)\| + P \leq r, \ \forall \ k \in M, \ (3.2) \text{ and } (3.3) \text{ hold.} \tag{3.12}$$
B) The \textit{ad-type} optimisation problem from (2.9) becomes
\[
\min_{(y, P, t) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m} \left\{ \sum_{k \in M} \left( p_k^T (x - y) + bt_k \right) + P \right\} \\
\text{s.t. } \|Q_k(x - y)\| \leq t_k, \; \forall \; k \in M, \; (3.2) \; \text{and} \; (3.3) \; \text{hold.}
\]

C) The \textit{wa-type} optimisation problem from (2.10) becomes
\[
\min_{(y, P, t) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m} \left\{ \sum_{k \in M} \lambda_k \left( p_k^T (x - y) + bt_k \right) + P \right\} \\
\text{s.t. } \|Q_k(x - y)\| \leq t_k, \; \forall \; k \in M, \; (3.2) \; \text{and} \; (3.3) \; \text{hold.}
\]

D) The \textit{wwc-type} optimisation problem from (2.11) becomes
\[
\min_{(y, P, r, s, u) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^m} \{ r + P \} \\
\text{s.t. } s + \frac{1}{l} 1^T u \leq r, \; 0 \leq u, \; (3.2) \; \text{and} \; (3.3) \; \text{hold,}
\]
\[
p_k^T (x - y) + b\|Q_k(x - y)\| - s \leq u_k, \; \forall \; k \in M.
\]

3.2. \textbf{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 \((\bar{R}, \bar{P})\) is also Pareto optimal if and only if there exists no other feasible solution \((\tilde{R}, \tilde{P})\) such that
\[
\rho(X - \tilde{R}[X]; P_k) + \tilde{P} \leq \rho(X - \bar{R}^*[X]; P_k) + P^* \; \forall \; k \in M,
\]
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 \((\bar{R}^*, \bar{P}^*)\) are also Pareto optimal. That is, the solutions of the Additive Model (2.9) and Weighted Average Model (2.10) with strictly positive \(\lambda_k\)'s (for all \(k \in M\)) 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 \textit{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 \textit{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 \( f_k : A \rightarrow \mathbb{R}, g_k : A \rightarrow \mathbb{R}^{n_k} \) be some functions over a convex set \( A \), 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.

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

with \( \sigma \) being a permutation of \( \mathcal{M} \). The optimisation problem becomes:

\[
\min_{x \in A} \sum_{i=1}^{l} \lambda_i f^{(i)}(x), \quad \text{s.t} \quad g_k(x) \in A_k, \quad \forall \ k \in \mathcal{M},
\]

where \( \lambda_k \)'s are positive scalars and \( A_k \) 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 \( A \). Using the notation from Problem 3.1, a feasible solution \( x^* \), i.e. \( g_k(x^*) \in A_k \) for all \( k \in \mathcal{M} \), is Pareto optimal if there is no other feasible solution \( y \), i.e. \( g_k(y) \in A_k \) for all \( k \in \mathcal{M} \), such that \( f_k(y) \leq f_k(x^*) \) 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 (3.16) that are Pareto optimal as well without massively increase the computational effort.

**Theorem 3.1.** Let \( x^* \) be an optimal solution of (3.16). Then, \( x^* \) is also Pareto optimal if the optimal objective function value of the following optimisation problem

\[
\min_{y \in A} \sum_{k \in \mathcal{M}} (f_k(y) - f_k(x^*)), \quad \text{s.t} \quad g_k(y) \in A_k, \quad f_k(y) - f_k(x^*) \leq 0, \quad \forall \ k \in \mathcal{M}.
\]

is zero. On the other hand, if the optimal value of (3.17) is negative, then any optimal solution \( y^* \) of (3.17) solves (3.16) 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 \( x = (x_1, x_2, \ldots, x_n)^T \) to be drawn from the underlying distribution of \( X \). Without loss of generality, we further assume that \( X \) is LogNormal distributed with mean \( E(X) = 5,000 \) and standard deviation \( \sqrt{3} \times E(X) \). The premium principle \( \Phi \) 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

\(^2\)A set \( B \) is a convex cone if and only if for any scalars \( a, b > 0 \), \( ax + by \in B \) given that \( x, y \in B \).
insurance cost is \( P = \frac{(1+\theta)E(X)}{2} \). 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: 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: Inverse Gaussian distribution with parameters \((\mu, \sigma)\) and cdf \( F(z) = \Phi\left(\sqrt{\frac{2}{\nu}}(\frac{x}{\mu} - 1)\right) + \Phi\left(-\sqrt{\frac{2}{\nu}}(\frac{x}{\mu} + 1)\right)e^{2\lambda/\mu}, z > 0. \)

For implementation purposes, we should define the probability vector \( p_k \) for all \( k \in M \) by discriminating the Maximum Likelihood estimated model with the sample observation \( x \). That is,

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

where by convention \( x_0 = -\infty \) and \( x_{n+1} = \infty \). Moreover, \( \hat{\nu} \) is the Maximum Likelihood Estimate based on the sample \( x \). Let us also denote the true underlying distribution of \( X \) and its corresponding probability vector as Model 0 and \( p_0 \), respectively. Then, \( p_0 \) 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: \( M_5 := \{1, 2, 3, 4, 5\}, \ M_4 := \{1, 3, 4, 5\}, \ M_2 := \{1, 5\}, \ M_1^* := \{2, 3, 4, 5\} \) and \( M_5^* := \{2, 5\} \). Note that the underlying distribution of \( X \) is LogNormal, and thus, we have deliberately excluded Model 2 from \( M_2 \) and \( M_4 \) in order to investigate the impact of model misidentification, when the “true” model is discarded.

We also need to specify the weights \( \lambda_k \)'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, \( AIC_{min} := \min_{k \in M} AIC_k \).

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

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

Also, if we denote \( k^* \) such that \( AIC_{min} = AIC_{k^*} \), then \( RL_{k^*} = 1 \) and \( \lambda_{k^*} \geq \lambda_k \) for all \( k \in M \). 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 \( (y_{wc}^*, P_{wc}^*), (y_{ad}^*, P_{ad}^*), (y_{wa}^*, P_{wa}^*) \) and \( (y_{wuc}^*, P_{wuc}^*) \), respectively. In particular, \( y_{wc}^* \) 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 $(y^*_T, P^*_T)$. In fact, $(y^*_T, P^*_T)$ could be obtained by solving (2.9) with $M = \{0\}$. The robustness of a generic optimal solution $y^*$ 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 $y^*_A$ and $y^*_B$, 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 $(y^*_{AIC}, P^*_{AIC})$. The second “non-robust” model is called the Elicitable Model and its solution is denoted as $(y^*_e, P^*_e)$. 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. $k^* := \arg \min_{k \in M} AIC_k$. Then, $(y^*_{AIC}, P^*_{AIC})$ is found by solving (2.9) with $M = k^*$.

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

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

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

The functional $f$ is elicitable relative to a class of probability measure $\Pi$ if and only if there exists a scoring function $S$ that is strictly consistent for $f$ relative to $\Pi$. The concept of elicitability 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\(_{\alpha}\) is elicitable and
\[
\mathbb{E}_P S_g(\text{VaR}_\alpha(X; P), x) \leq \mathbb{E}_P S_g(y, x)
\]  
(4.2)
for any real number \(y \in \mathbb{R}\), where \(S_g(t, x) = (I\{t \geq x\} - \alpha) (g(t) - g(x))\) 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 \(P_k\) becomes
\[
\sum_{i=1}^{n} p_{ik} S_g(x_{p(k)}, x_i) \leq \sum_{i=1}^{n} p_{ik} S_g(y, x_i).
\]
As a result, whenever the “true” probability distribution \(P_k\) is unknown, but \(m\) probability candidate models are available, one may choose the “best” distribution \(k^*\) that gives the lowest expected score, i.e.
\[
k^* = \arg \min_k \sum_{i=1}^{n} p_{ik} S(x_{p(k)}, x_i),
\]
and hence, the “best” estimate of \(\text{VaR}_\alpha\) is \(x_{p(k^*)}\). Finally, the non-robust optimal elicitation solution \((y^*, P^*_e)\) may be found by solving the following LP for all \(l \in M\):
\[
\begin{align*}
\min_{(y,P) \in \mathbb{R}^n \times \mathbb{R}} \{ & x_{p(l)} - y_{p(l)} + P \\ s.t. \quad & \sum_{i=1}^{n} p_{il} S(x_{p(l)} - y_{p(l)}, x_i - y_i) \leq \sum_{i=1}^{n} p_{il} S(x_{p(l)} - y_{p(l)}, x_i - y_i), \quad \forall k \in M, \\ & (1 + \theta) \mathbf{p}_k^T \mathbf{y} \leq P \leq \overline{P}, \quad \forall k \in M, \\ & 0 \leq \mathbf{y} \leq \mathbf{x}, 0 \leq A\mathbf{y} \leq A\mathbf{x}.
\end{align*}
\]  
(4.3)
Let \((y^*_{el}, P^*_{el})\) be the optimal solution found for the above LP under distribution \(l\) and let \(l^*\) be the probability model choice under the elicitation criterion, which is given by the one with the lowest objective function (4.3) amongst all \(l \in M\). Therefore, the Elicitability Model optimal solution is \((y^*_{e}, P^*_{e}) := (y^*_{el}, P^*_{el})\). 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 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. We first investigate the results for the VaR\(_{\alpha}\)-based optimisation problems when \(\alpha = 0.75\), which are illustrated in Tables 4.1–4.3. Our numerical 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 (3.6) is compared to the other three robust models (3.4), (3.5)
Table 4.1. Results when (3.6) is compared to (3.4), (3.5) and the AIC model for the VaR_{0.75}-based solutions under various sample sizes \(n\) and collections of candidate models \(\{M_2, M_4, M_5\}\).

<table>
<thead>
<tr>
<th></th>
<th>(n = 25)</th>
<th>(n = 50)</th>
<th>(n = 100)</th>
<th>(n = 250)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M_5)</td>
<td>(M_4)</td>
<td>(M_2)</td>
<td>(M_5)</td>
</tr>
<tr>
<td>(S^\ast_{wc} \succ S^\ast_{wa})</td>
<td>215</td>
<td>224</td>
<td>226</td>
<td>184</td>
</tr>
<tr>
<td>(S^\ast_{wa} \succ S^\ast_{wc})</td>
<td>285</td>
<td>276</td>
<td>273</td>
<td>316</td>
</tr>
<tr>
<td>(S^\ast_{ad} \succ S^\ast_{wa})</td>
<td>196</td>
<td>166</td>
<td>134</td>
<td>151</td>
</tr>
<tr>
<td>(S^\ast_{wa} \succ S^\ast_{ad})</td>
<td>304</td>
<td>334</td>
<td>366</td>
<td>349</td>
</tr>
<tr>
<td>(S^\ast_{wa} \succ S^\ast_{AIC})</td>
<td>224</td>
<td>216</td>
<td>221</td>
<td>198</td>
</tr>
<tr>
<td>(S^\ast_{AIC} \succ S^\ast_{wa})</td>
<td>275</td>
<td>284</td>
<td>278</td>
<td>302</td>
</tr>
<tr>
<td>(S^\ast_{wa} \succ S^\ast_{e})</td>
<td>320</td>
<td>352</td>
<td>349</td>
<td>386</td>
</tr>
<tr>
<td>(S^\ast_{e} \succ S^\ast_{wa})</td>
<td>180</td>
<td>148</td>
<td>149</td>
<td>114</td>
</tr>
</tbody>
</table>

Table 4.2. Results when (3.6) is compared to (3.4), (3.5) and the AIC model for the VaR_{0.75}-based solutions under various sample sizes \(n\) and collections of candidate models \(\{M^\ast_2, M^\ast_4, M_5\}\).

<table>
<thead>
<tr>
<th></th>
<th>(n = 25)</th>
<th>(n = 50)</th>
<th>(n = 100)</th>
<th>(n = 250)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M_5)</td>
<td>(M^\ast_4)</td>
<td>(M^\ast_2)</td>
<td>(M_5)</td>
</tr>
<tr>
<td>(S^\ast_{wc} \succ S^\ast_{wa})</td>
<td>215</td>
<td>211</td>
<td>246</td>
<td>184</td>
</tr>
<tr>
<td>(S^\ast_{wa} \succ S^\ast_{wc})</td>
<td>285</td>
<td>289</td>
<td>254</td>
<td>316</td>
</tr>
<tr>
<td>(S^\ast_{ad} \succ S^\ast_{wa})</td>
<td>196</td>
<td>152</td>
<td>133</td>
<td>151</td>
</tr>
<tr>
<td>(S^\ast_{wa} \succ S^\ast_{ad})</td>
<td>304</td>
<td>348</td>
<td>367</td>
<td>349</td>
</tr>
<tr>
<td>(S^\ast_{wa} \succ S^\ast_{AIC})</td>
<td>224</td>
<td>177</td>
<td>236</td>
<td>198</td>
</tr>
<tr>
<td>(S^\ast_{AIC} \succ S^\ast_{wa})</td>
<td>275</td>
<td>323</td>
<td>263</td>
<td>302</td>
</tr>
<tr>
<td>(S^\ast_{wa} \succ S^\ast_{e})</td>
<td>320</td>
<td>291</td>
<td>275</td>
<td>386</td>
</tr>
<tr>
<td>(S^\ast_{e} \succ S^\ast_{wa})</td>
<td>180</td>
<td>209</td>
<td>223</td>
<td>114</td>
</tr>
</tbody>
</table>

and (3.7), respectively. Recall that when \(l = 1\), the Weighted Worst-case Model (3.7) becomes the classic Worst-case model (3.4), and thus, we only solve (3.7) under \(M = M_5, M_4, M_2\) with \(2 \leq l \leq m - 1\). Note that the \(l = 4\) case only exists when \(M = M_5\). We noticed that the Weighted Average Model stands as the most robust model in all comparisons, especially when
Table 4.3. Comparison between the VaR$_{0.75}$-based solutions of (3.6) and (3.7) for various sample sizes $n$ and collections of candidate models $\{M_5, M_4, M^*_4\}$.

<table>
<thead>
<tr>
<th></th>
<th>$n = 25$</th>
<th></th>
<th>$n = 50$</th>
<th></th>
<th>$n = 100$</th>
<th></th>
<th>$n = 250$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M_5$</td>
<td>$M_4$</td>
<td>$M^*_4$</td>
<td>$M_5$</td>
<td>$M_4$</td>
<td>$M^*_4$</td>
<td>$M_5$</td>
<td>$M_4$</td>
</tr>
<tr>
<td>$S_{wuc}^* &gt; S_{wa}^* (l=2)$</td>
<td>202</td>
<td>201</td>
<td>204</td>
<td>171</td>
<td>182</td>
<td>160</td>
<td>123</td>
<td>116</td>
</tr>
<tr>
<td>$S_{wa}^* &gt; S_{wuc}^* (l=2)$</td>
<td>298</td>
<td>299</td>
<td>296</td>
<td>329</td>
<td>317</td>
<td>340</td>
<td>377</td>
<td>384</td>
</tr>
<tr>
<td>$S_{wuc}^* &gt; S_{wa}^* (l=3)$</td>
<td>238</td>
<td>227</td>
<td>259</td>
<td>182</td>
<td>202</td>
<td>219</td>
<td>128</td>
<td>136</td>
</tr>
<tr>
<td>$S_{wa}^* &gt; S_{wuc}^* (l=3)$</td>
<td>262</td>
<td>273</td>
<td>241</td>
<td>318</td>
<td>298</td>
<td>281</td>
<td>372</td>
<td>364</td>
</tr>
<tr>
<td>$S_{wuc}^* &gt; S_{wa}^* (l=4)$</td>
<td>241</td>
<td></td>
<td>203</td>
<td>153</td>
<td></td>
<td>105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{wa}^* &gt; S_{wuc}^* (l=4)$</td>
<td>259</td>
<td></td>
<td>297</td>
<td>347</td>
<td></td>
<td>395</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 (3.6) 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 (3.6) does outperform the elicitation criterion. However, the performance of the 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 $\Delta^*$ 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 $\Delta^*$'s might be informative as well. Figure 4.1 compares the boxplots between $\Delta_{wa}^*$ and $\Delta_{AIC}^*$. In each of the boxplots, the median of $\Delta^*$'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 $\Delta_{wa}^*$ and $\Delta_{AIC}^*$ shrinks dramatically when the sample size $n$ grows for all distribution collections $\mathcal{M} \in \{\mathcal{M}_5, \mathcal{M}_4, \mathcal{M}_2, \mathcal{M}_4^*\}$. 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 (3.4)–(3.10) in the VaR-based case, Figure 4.1 shows that $\Delta_{wa}^*$ and $\Delta_{AIC}^*$ 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 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 (3.8)–(3.11). Tables 4.4–4.6 have shown a similar pattern as seen
Figure 4.1. Boxplots comparing $\Delta^*_w$ and $\Delta^*_AIC$ computed from the 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 $\Delta^*_w/\Delta^*_AIC$. The top row boxplots are corresponding to distribution collections $M_5$, $M_4$ and $M_2$, while the bottom row relates to $M_4^*$ and $M_2^*$, respectively.

<table>
<thead>
<tr>
<th></th>
<th>$n = 25$</th>
<th>$n = 50$</th>
<th>$n = 100$</th>
<th>$n = 250$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S^<em>_{wc} &gt; S^</em>_w$</td>
<td>189</td>
<td>198</td>
<td>180</td>
<td>155</td>
</tr>
<tr>
<td>$S^<em>_{wc} &gt; S^</em>_w$</td>
<td>311</td>
<td>300</td>
<td>320</td>
<td>345</td>
</tr>
<tr>
<td>$S^<em>_{ad} &gt; S^</em>_w$</td>
<td>188</td>
<td>231</td>
<td>218</td>
<td>199</td>
</tr>
<tr>
<td>$S^<em>_{ad} &gt; S^</em>_w$</td>
<td>312</td>
<td>269</td>
<td>269</td>
<td>265</td>
</tr>
<tr>
<td>$S^<em>_{wa} &gt; S^</em>_{AIC}$</td>
<td>276</td>
<td>298</td>
<td>271</td>
<td>253</td>
</tr>
<tr>
<td>$S^<em>_{wa} &gt; S^</em>_{AIC}$</td>
<td>224</td>
<td>202</td>
<td>229</td>
<td>247</td>
</tr>
</tbody>
</table>

Table 4.4. Results when (3.10) is compared to (3.8), (3.9) and the AIC model for the CVaR$_{0.75}$-based solutions under various sample sizes $n$ and collections of candidate models $\{M_2, M_4, M_5\}$.
Table 4.5. Results when (3.10) is compared to (3.8), (3.9) and the AIC model for the CVaR_{0.75}-based solutions under various sample sizes \( n \) and collections of candidate models \( \{M_5, M_4, M_2\} \).

<table>
<thead>
<tr>
<th></th>
<th>( n = 25 )</th>
<th>( n = 50 )</th>
<th>( n = 100 )</th>
<th>( n = 250 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M_5 )</td>
<td>( M_4 )</td>
<td>( M_5 )</td>
<td>( M_4 )</td>
</tr>
<tr>
<td>( S^<em><em>{w</em>{wc}} \succ S^</em><em>{w</em>{wa}} ) ( (l=2) )</td>
<td>189</td>
<td>197</td>
<td>221</td>
<td>198</td>
</tr>
<tr>
<td>( S^<em>_{wa} \succ S^</em><em>{w</em>{wc}} ) ( (l=2) )</td>
<td>311</td>
<td>303</td>
<td>279</td>
<td>300</td>
</tr>
<tr>
<td>( S^<em>_{ad} \succ S^</em>_{wa} ) ( (l=3) )</td>
<td>188</td>
<td>193</td>
<td>197</td>
<td>231</td>
</tr>
<tr>
<td>( S^<em>_{wa} \succ S^</em>_{ad} ) ( (l=3) )</td>
<td>312</td>
<td>307</td>
<td>303</td>
<td>269</td>
</tr>
<tr>
<td>( S^<em>_{wa} \succ S^</em>_{AIC} ) ( (l=3) )</td>
<td>276</td>
<td>281</td>
<td>286</td>
<td>298</td>
</tr>
<tr>
<td>( S^<em>_{AIC} \succ S^</em>_{wa} ) ( (l=3) )</td>
<td>224</td>
<td>219</td>
<td>213</td>
<td>202</td>
</tr>
</tbody>
</table>

Table 4.6. Comparison between the CVaR_{0.75}-based solutions of (3.10) and (3.11) for various sample sizes \( n \) and collections of candidate models \( \{M_5, M_4, M_2\} \).

<table>
<thead>
<tr>
<th></th>
<th>( n = 25 )</th>
<th>( n = 50 )</th>
<th>( n = 100 )</th>
<th>( n = 250 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M_5 )</td>
<td>( M_4 )</td>
<td>( M_5 )</td>
<td>( M_4 )</td>
</tr>
<tr>
<td>( S^<em><em>{w</em>{wc}} \succ S^</em><em>{w</em>{wa}} ) ( (l=2) )</td>
<td>199</td>
<td>268</td>
<td>203</td>
<td>204</td>
</tr>
<tr>
<td>( S^<em>_{wa} \succ S^</em><em>{w</em>{wc}} ) ( (l=2) )</td>
<td>301</td>
<td>232</td>
<td>297</td>
<td>296</td>
</tr>
<tr>
<td>( S^<em><em>{w</em>{wc}} \succ S^</em>_{wa} ) ( (l=3) )</td>
<td>233</td>
<td>243</td>
<td>248</td>
<td>225</td>
</tr>
<tr>
<td>( S^<em>_{wa} \succ S^</em><em>{w</em>{wc}} ) ( (l=3) )</td>
<td>266</td>
<td>257</td>
<td>252</td>
<td>274</td>
</tr>
<tr>
<td>( S^<em><em>{w</em>{wc}} \succ S^</em>_{wa} ) ( (l=4) )</td>
<td>235</td>
<td>242</td>
<td>253</td>
<td>253</td>
</tr>
<tr>
<td>( S^<em>_{wa} \succ S^</em><em>{w</em>{wc}} ) ( (l=4) )</td>
<td>265</td>
<td>258</td>
<td>247</td>
<td>344</td>
</tr>
</tbody>
</table>

in the VaR case, where the optimal solutions found under the Weighted Average Model (3.10) turn out to be the most robust among the four models (3.8)–(3.11). 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 sizes \( n \) and distribution collections \( M \). Boxplots are also produced to better compare \( \Delta_{wa}^* \) and \( \Delta_{AIC}^* \) for CVaR-based optimisations, which could be found in Figure 4.2. Although the median value of \( \Delta_{wa}^* \) and \( \Delta_{AIC}^* \) are very similar under various sample sizes and distribution collections, the range of \( \Delta_{AIC}^* \) is in general larger than that of \( \Delta_{wa}^* \), especially when the sample is small. Therefore, the overall
Figure 4.2. Boxplots comparing $\Delta^*_w\text{ and }\Delta^*_AIC$ computed from the 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 $\Delta^*_w/\Delta^*_AIC$. The top row boxplots are corresponding to distribution collections $M_5$, $M_4$ and $M_2$, while the bottom row relates to $M^*_4$ and $M^*_2$, respectively.

Evidence tells us that our Weighted Average Model (3.10) 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 (3.8)–(3.11) 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 is small. The last four rows displayed in Tables 4.7 and 4.8 summarise comparisons amongst optimal contracts found under $PHT$-based criterion with $\alpha = 0.9$ and 0.2. The performance of the Weighted Average Model (3.10) 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 the greater this value is, the more risk aversion the decision-maker is. When $\alpha$ 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. $\alpha$ 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
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 \(\{M_2, M_4, M_5\}\).

<table>
<thead>
<tr>
<th></th>
<th>(n = 25)</th>
<th>(n = 50)</th>
<th>(n = 100)</th>
<th>(n = 250)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M_5)</td>
<td>(M_4)</td>
<td>(M_2)</td>
<td>(M_5)</td>
<td>(M_4)</td>
</tr>
<tr>
<td>(S_{wa}^* \succ S_{wc}^* (\alpha = 0.9))</td>
<td>204 144 208</td>
<td>247 211 210</td>
<td>224 201 138</td>
<td>247 229 80</td>
</tr>
<tr>
<td>(S_{wa}^* \succ S_{wc}^* (\alpha = 0.9))</td>
<td>296 356 292</td>
<td>276 289 362</td>
<td>253 271 120</td>
<td></td>
</tr>
<tr>
<td>(S_{ad}^* \succ S_{wa}^* (\alpha = 0.9))</td>
<td>74 65 49</td>
<td>49 30 11</td>
<td>27 19 1</td>
<td></td>
</tr>
<tr>
<td>(S_{wa}^* \succ S_{ad}^* (\alpha = 0.9))</td>
<td>426 435 451</td>
<td>451 470 489</td>
<td>473 481 499</td>
<td></td>
</tr>
<tr>
<td>(S_{wa}^* \succ S_{AIC}^* (\alpha = 0.9))</td>
<td>114 142 153</td>
<td>24 88 165</td>
<td>3 92 200</td>
<td></td>
</tr>
<tr>
<td>(S_{AIC}^* \succ S_{wa} (\alpha = 0.9))</td>
<td>386 358 347</td>
<td>476 412 335</td>
<td>497 408 300</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(n = 25)</th>
<th>(n = 50)</th>
<th>(n = 100)</th>
<th>(n = 250)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M_5)</td>
<td>(M_4)</td>
<td>(M_2)</td>
<td>(M_5)</td>
<td>(M_4)</td>
</tr>
<tr>
<td>(S_{wa}^* \succ S_{AIC}^* (\alpha = 0.2))</td>
<td>235 235 243</td>
<td>267 286 284</td>
<td>229 254 254</td>
<td>210 223 251</td>
</tr>
<tr>
<td>(S_{AIC}^* \succ S_{wa}^* (\alpha = 0.2))</td>
<td>264 265 254</td>
<td>233 214 216</td>
<td>271 246 246</td>
<td>290 277 249</td>
</tr>
</tbody>
</table>

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 \(\{M_2, M_4, M_5\}\).

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 \(\Delta_{wa}^*\) and \(\Delta_{AIC}^*\) for the PHT-based case with \(c = 0.2\). As before, the range of \(\Delta_{wa}^*\) and \(\Delta_{AIC}^*\) are very similar in most of the comparisons, especially
Table 4.9. Comparison between the PHT-based ($\alpha = 0.9$) solutions for various sample sizes $n$ and collection of candidate models $\{M_5, M_4, M^*_4\}$.

<table>
<thead>
<tr>
<th></th>
<th>$n = 25$</th>
<th>$n = 50$</th>
<th>$n = 100$</th>
<th>$n = 250$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M_5$</td>
<td>$M_4$</td>
<td>$M^*_4$</td>
<td>$M_5$</td>
</tr>
<tr>
<td>$S^<em>_{wwc} \succ S^</em>_{wa} (l=2)$</td>
<td>172</td>
<td>179</td>
<td>196</td>
<td>202</td>
</tr>
<tr>
<td>$S^<em>_{wa} \succ S^</em>_{wwc} (l=2)$</td>
<td>328</td>
<td>321</td>
<td>304</td>
<td>298</td>
</tr>
<tr>
<td>$S^<em>_{wwc} \succ S^</em>_{wa} (l=3)$</td>
<td>182</td>
<td>191</td>
<td>148</td>
<td>182</td>
</tr>
<tr>
<td>$S^<em>_{wa} \succ S^</em>_{wwc} (l=3)$</td>
<td>318</td>
<td>309</td>
<td>352</td>
<td>318</td>
</tr>
<tr>
<td>$S^<em>_{wwc} \succ S^</em>_{wa} (l=4)$</td>
<td>161</td>
<td>174</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>$S^<em>_{wa} \succ S^</em>_{wwc} (l=4)$</td>
<td>339</td>
<td>326</td>
<td>405</td>
<td>368</td>
</tr>
</tbody>
</table>

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 (3.10).

Figure 4.3. Boxplots comparing $\Delta^*_{wa}$ and $\Delta^*_{AIC}$ 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 $M_5, M_4$ and $M_2$, while the bottom row relates to $M^*_4$ and $M^*_2$, respectively.
The last set of results of the section considers the robustness of the SD-based optimal contracts where \( b = 0.5 \). Table 4.10 compares the Weighted Average Model to the Weight Worst-case Model 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. Additional comparisons between other robust methods are not displayed, since the results are quite similar, i.e. there is no evidence to believe that the Worst-case or Additive Models perform any better than the Weighted Average Model. Further, the Weighted Worst-case Model is compared to the “non-robust” AIC Model in Table 4.11. One may find that our robust Weighted Worst-case model is recommended over the “non-robust” AIC Model when the sample size is relatively small, e.g. \( n = 25, 50 \), otherwise the AIC Model is preferred. 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.

<table>
<thead>
<tr>
<th></th>
<th>( n = 25 )</th>
<th>( n = 50 )</th>
<th>( n = 100 )</th>
<th>( n = 250 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_{\text{wwc}}^* \succ S_{\text{wa}}^* ) (( l = 2 ))</td>
<td>296</td>
<td>260</td>
<td>298</td>
<td>294</td>
</tr>
<tr>
<td>( S_{\text{wa}}^* \succ S_{\text{wwc}}^* ) (( l = 2 ))</td>
<td>204</td>
<td>240</td>
<td>202</td>
<td>174</td>
</tr>
<tr>
<td>( S_{\text{wwc}}^* \succ S_{\text{wa}}^* ) (( l = 3 ))</td>
<td>309</td>
<td>319</td>
<td>308</td>
<td>347</td>
</tr>
<tr>
<td>( S_{\text{wa}}^* \succ S_{\text{wwc}}^* ) (( l = 3 ))</td>
<td>191</td>
<td>181</td>
<td>192</td>
<td>153</td>
</tr>
<tr>
<td>( S_{\text{wwc}}^* \succ S_{\text{wa}}^* ) (( l = 4 ))</td>
<td>326</td>
<td>350</td>
<td>192</td>
<td>150</td>
</tr>
<tr>
<td>( S_{\text{wa}}^* \succ S_{\text{wwc}}^* ) (( l = 4 ))</td>
<td>174</td>
<td>174</td>
<td>192</td>
<td>150</td>
</tr>
</tbody>
</table>

Table 4.10. Comparison between the SD-based (\( b=0.5 \)) solutions of (3.14) and (3.15) for various sample sizes \( n \) and collection of candidate models \( \{ M_5, M_4, M_4^* \} \).

It is also worth mentioning as a final remark that if we compare all the boxplots in Figures 4.1–4.4, the \( \Delta^* \) 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.
### Table 4.11

Comparison between the $SD$-based ($b=0.5$) solutions of (3.15) and the non-robust AIC model for various sample sizes $n$ and collection of candidate models $\{M_5, M_4, M_4^*, M_5^*\}$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$M_5$</th>
<th>$M_4$</th>
<th>$M_4^*$</th>
<th>$M_5$</th>
<th>$M_4$</th>
<th>$M_4^*$</th>
<th>$M_5$</th>
<th>$M_4$</th>
<th>$M_4^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>270</td>
<td>277</td>
<td>278</td>
<td>281</td>
<td>302</td>
<td>287</td>
<td>238</td>
<td>240</td>
<td>254</td>
</tr>
<tr>
<td>50</td>
<td>230</td>
<td>223</td>
<td>222</td>
<td>219</td>
<td>198</td>
<td>213</td>
<td>262</td>
<td>260</td>
<td>246</td>
</tr>
<tr>
<td>100</td>
<td>282</td>
<td>292</td>
<td>279</td>
<td>289</td>
<td>312</td>
<td>290</td>
<td>237</td>
<td>245</td>
<td>249</td>
</tr>
<tr>
<td>250</td>
<td>218</td>
<td>208</td>
<td>221</td>
<td>211</td>
<td>188</td>
<td>210</td>
<td>263</td>
<td>255</td>
<td>251</td>
</tr>
</tbody>
</table>

**Figure 4.4.** Boxplots comparing $\Delta_{wa}^*$ and $\Delta_{AIC}^*$ computed from the $SD$-based ($b = 0.2$) 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 $M_5, M_4$ and $M_2$, while the bottom row relates to $M_4^*$ and $M_5^*$, respectively.
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 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.

6. Proofs

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

\[
\min_{(t, s) \in A \times \mathbb{R}} \left\{ s + \frac{1}{l} \sum_{i=1}^{m} \left( f(t; \omega_i) - s \right)_+ \right\},
\]

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 \((t^*, s^*, u^*)\). It is noticed that the objective function in (2.7) is increasing in \( u_i \) for all \( i \in \mathcal{M} \), and therefore, constraints \( f(t; \omega_i) \leq s + u_i \) and \( 0 \leq u \) ensure that \( u^*_i = \left( f(t^*; \omega_i) - s^* \right)_+ \) for all \( i \in \mathcal{M} \). Consequently, \((t^*, s^*)\) is also feasible to the problem (6.1). Suppose that \((t^*, s^*)\) is not the optimal solution to (6.1), then there must exist another feasible solution \((t', s')\) such that

\[
s' + \frac{1}{l} \sum_{i=1}^{m} \left( f(t'; \omega_i) - s' \right)_+ < s^* + \frac{1}{l} \sum_{i=1}^{m} \left( f(t^*; \omega_i) - s^* \right)_+ = s^* + \frac{1}{l} \sum_{i=1}^{m} u^*_i.
\]

Note that \((t', s', u')\) with \( u'_i = \left( f(t'; \omega_i) - s' \right)_+ \) 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 \((t^*, s^*, u^*)\) 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 \((t^*, s^*)\) is the optimal solution to (6.1). Then, \((t^*, s^*, u^*)\) with \(u^*_i = \left( f(t^*; \omega_i) - s^* \right)_+ \) for all \(i \in M\) is also feasible to (2.7). If \((t^*, s^*, u^*)\) is not an optimal solution to (2.7), there must exist another feasible solution \((t', s', u')\) such that

\[
s' + \frac{1}{l} \sum_{i=1}^m u'_i < s^* + \frac{1}{l} \sum_{i=1}^m u^*_i = s^* + \frac{1}{l} \sum_{i=1}^m \left( f(t^*; \omega_i) - s^* \right)_+. \tag{6.3}
\]

Since the constraints \(f(t; \omega_i) \leq s + u_i\) and \(0 \leq u\) in (2.7) will ensure \(u'_i = \left( f(t'; \omega_i) - s' \right)_+\), \((t', s')\) 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 \((t^*, s^*)\) 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. \(\square\)

**Proof of Theorem 3.1.** Let us first show that an optimal solution \(x^*\) of (3.16) must be Pareto optimal when the optimal objective function value in (3.17) is zero. If \(x^*\) is not Pareto optimal, then there must exist another feasible solution \(\hat{y}\) of (3.16) such that \(f_k(\hat{y}) \leq f_k(x^*)\) for all \(k \in M\) with at least one inequality sign being strict. Thus, \(\hat{y}\) is feasible in (3.17) and

\[
\sum_{k \in M} \left( f_k(\hat{y}) - f_k(x^*) \right) < 0,
\]

which contradicts the statement that the optimal objective function value of (3.17) is zero. Thus, \(x^*\) must be Pareto optimal.

Next, we show that when the optimal objective function value of (3.17) is negative, any optimal solution \(y^*\) of (3.17) solves (3.16) as well and is Pareto optimal. Now, \(f_k(y^*) \leq f_k(x^*)\) for any \(k \in M\), since \(y^*\) is feasible in (3.17), which in turn gives that \(f^{(k)}(y^*) \leq f^{(k)}(x^*)\) for any \(k \in M\). The latter and the fact \(\lambda_k's\) are positive imply that \(y^*\) must solve (3.16), since \(x^*\) solves (3.16). Assume now that \(y^*\) is not Pareto optimal. Therefore, there must exist another feasible solution \(\hat{y}\) of (3.16) such that \(f_k(\hat{y}) \leq f_k(y^*)\) for all \(k \in M\) with at least one inequality sign being strict. Consequently, \(\hat{y}\) is feasible in (3.17) and

\[
\sum_{k \in M} \left( f_k(\hat{y}) - f_k(x^*) \right) < \sum_{k \in M} \left( f_k(y^*) - f_k(x^*) \right),
\]

which contradicts the fact that \(y^*\) is an optimal solution of (3.17). Therefore, \(y^*\) must be Pareto optimal. The proof is now complete. \(\square\)
References


