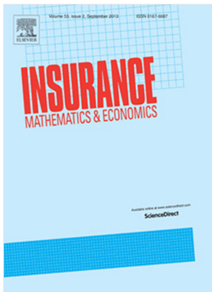
View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by International Institute for Applied Systems Analysis (IIASA)

Accepted Manuscript



Optimal XL-insurance under Wasserstein-type ambiguity

Corina Birghila, Georg Ch. Pflug

PII: S0167-6687(18)30335-4
DOI: https://doi.org/10.1016/j.insmatheco.2019.05.005
Reference: INSUMA 2561

To appear in: *Insurance: Mathematics and Economics*

Received date : 9 August 2018
Revised date : 23 May 2019
Accepted date : 24 May 2019

Please cite this article as: C. Birghila and G.C. Pflug, Optimal XL-insurance under Wasserstein-type ambiguity. *Insurance: Mathematics and Economics* (2019), https://doi.org/10.1016/j.insmatheco.2019.05.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Optimal XL-insurance under Wasserstein-type ambiguity

Corina Birghila *[1] and Georg Ch. Pflug †‡[1,2]

[1]ISOR, University of Vienna, Oskar Morgenstern-Platz 1, A-1090, Vienna, Austria
[2]IIASA, International Institute for Applied Systems Analysis, Laxenburg, Austria

Abstract

We study the problem of optimal insurance contract design for risk management under a budget constraint. The contract holder takes into consideration that the loss distribution is not entirely known and therefore faces an ambiguity problem. For a given set of models, we formulate a minimax optimization problem of finding an optimal insurance contract that minimizes the distortion risk functional of the retained loss with premium limitation. We demonstrate that under the average value-at-risk measure, the entrance-excess of loss contracts are optimal under ambiguity, and we solve the distributionally robust optimal contract-design problem. It is assumed that the insurance premium is calculated according to a given baseline loss distribution and that the ambiguity set of possible distributions forms a neighborhood of the baseline distribution. To this end, we introduce a *contorted Wasserstein distance*. This distance is finer in the tails of the distributions compared to the usual Wasserstein distance.

JEL code. G22, D81.
Keywords: insurance contract optimization, model error, minimax solution, distributional robustness.

1 Introduction

Finding an optimal insurance or reinsurance contract is an important topic in actuarial science, describing one of the most efficient tools for risk management. The works of Borch (1960) and Arrow (1963) were the first to discuss the structure of such contracts under budget constraints and with the risk quantified by variance or utility function. Since then, the problem of finding an optimal insurance contract has been studied under different market assumptions and under various risk preferences for the insurance participants. The expected utility framework analyzed in the aforementioned papers was further extended in the work of Raviv (1979),Young (1999) and Kaluszka (2001) among others. Another direction that drew substantial attention was the consideration of the optimal insurance contract that minimizes some risk functional, with the most common ones being the value-at-risk (V@R) and the average value-at-risk (AV@R). The problem was studied in Bernard and Tian (2009), Tan et al. (2011), Chi and Tan (2011), Chi and Tan (2013), Asia (2015) and Lo (2017a) under different choices of premium principle calculations.

The papers mentioned above rely on the assumption that the loss distribution is completely known. However, this assumption has been proven too restrictive. In most cases, approaches relying on such a hypothesis ignore possible errors in modeling, which can lead to an underestimation of the risk associated with the insured events. To overcome such drawbacks, we focus on the problem of quantifying the impact of model misspecification when designing insurance contracts. This issue becomes crucial in the context of extreme climatic events, where the need for more efficient insurance contracts has grown significantly in recent years.

The idea of considering model ambiguity has been used previously in environmental and finance applications to obtain more robust solutions. For instance, Zymler et al. (2013) used model ambiguity to control the probability that the water level in some reservoir remained within certain predefined limits. In portfolio optimization, we mention the work of Pflug and Wozabal (2007) and Esfahani and Kuhn (2017) as examples of constructing financial strategies when the underlying probability model is not completly

*(Corresponding author)corina.birghila@univie.ac.at
†georg.pflug@univie.ac.at
‡Declarations of interest: none.

ACCEPTED MANUSCRIPT

known. In actuarial science, there is a rich literature on worst-case risk measurement in the presence of incomplete probabilistic information, reviewed and extended in Goovaerts et al. (2011), but only in recent years a theoretical framework for the problem of optimal (re)insurance under ambiguity has been formulated (see Balbás et al. (2015), Asimit et al. (2017)).

To this end, it is important to mention that the terminology *ambiguity* was used before in literature to refer to the ambiguity averse attitude of market participants. More precisely, it was observed that people are more likely to gamble when the probability of loosing is known rather than when the loss probability is unknown, although the latter may be significantly lower. This paradox was first formulated by Daniel Ellsberg (Ellsberg (1961)) and is nowadays known as *Ellsberg's paradox*. The subsequent literature analyses the effect of ambiguity aversion on the structure of insurance contracts (see Klibanoff et al. (2005), Klibanoff et al. (2009), Alary et al. (2013)). Assuming that the ambiguous distribution of losses is parameterized by a finite set of priors, Gollier (2014) derives the optimal form of an insurance contact that maximizes the ex ante welfare of policyholder, under some insurance tariff constraints.

Our notion of ambiguity differs from the aforementioned Bayesian approach, since we do not assume any a priori structure on the probability models. The ambiguity arises from the uncertainty set of possible probabilistic models and leads to a minimax solution.

The objective of this paper is to incorporate ambiguity into the structure of the optimal insurance contract designed to protect against extreme natural events. In the context of low probability-high impact events, the climate-change dynamics and the scarcity of data could easily lead to model misspecification of the underlying loss distribution. These factors motivate use of the model ambiguity approach in the assessment and management of risk. The first objective of this paper is therefore to determine the structure of the optimal contract under model ambiguity. For a given set of models, we formulate a minimax optimization problem of finding an optimal insurance contract that minimizes the concave risk functional of the retained loss under the budget constraint of the premium. To compensate for possible model misspecification, the optimal decision is taken w.r.t. a set of non-parametric models. The ambiguity set is built using a modified version of the well-known Wasserstein distance, which is more sensitive to deviations in the tails of distributions. If the risk measure is the average value-at-risk, the optimization problem is solved using a distributionally robust optimization technique. We examine the dependence of the objective function as well as the parameters of the insurance contract on the tolerance level change. Numerical simulations illustrate the procedure.

The paper is organized as follows. Section 2 introduces the notions of risk measure and premium principle. As our focus is on low probability-high impact events insurance, we will provide a short introduction to extreme value theory (EVT), the statistical methodology used to model extreme events. In section 3 we specify the stochastic optimization problem of finding an optimal contract which is robust under a given set of models. The structure of the optimal solution is based on the Lagrange dual method for minimax optimization. In the next section, we consider the structure of the ambiguity set based on a modified version of the Wasserstein distance. The computational aspects of the minimax procedure are treated here. In section 5 we apply the framework described above to a dataset of tornado claims and study the impact of model ambiguity on the structure of an insurance contract.

2 Preliminaries and notations

Let $(\Omega, \mathcal{F}, P)$ be a probability space and L^1 be the set of all non-negative random variables X on Ω representing losses such that $\int_0^\infty |X(\omega)| \, dP(\omega) < \infty$.

Distortion risk measures. The distortion risk measure is defined using the notion of a distortion function.

Definition 2.1. A *(concave) distortion function* is a non-decreasing, concave function $g : [0, 1] \to [0, 1]$ such that $g(0) = 0$ and $g(1) = 1$.

Throughout the article we will focus on distortion risk measures built using concave distortion functions.

Definition 2.2. The *distortion risk measure* ρ^g of a random variable X with a distortion function g is

$$\rho^g(X) = \int_0^\infty g(1 - F(x)) \, dx, \tag{1}$$

where F is the distribution function of X.

ACCEPTED MANUSCRIPT

If X also takes negative values, then ρ^g is defined as

$$\rho^g(X) = \int_0^\infty g(1 - F(x))\,dx + \int_{-\infty}^0 [g(1 - F(x)) - 1]\,dx.$$

The definition of a distortion risk measure comes from the axiomatic characteristics of insurance pricing in Wang et al. (1997). The distortion risk measure ρ^g with concave distortion function g satisfies the following properties:

1. Properness: $\rho^g(X) \geq \mathbb{E}(X)$.
2. Positive homogeneity: $\rho^g(cX) = c\rho^g(X)$, for $c \in \mathbb{R}_+$.
3. Translation equivariance: $\rho^g(X + c) = \rho^g(X) + c$, for $c \in \mathbb{R}$.
4. Monotonicity: $\rho^g(X) \leq \rho^g(Y)$, for $X \leq Y$ a.s.
5. Comonotone additivity: $\rho^g(X + Y) = \rho^g(X) + \rho^g(Y)$, for comonotone random variables X, Y [1].
6. Version independence: $\rho^g(X) = \rho^g(Y)$, if $F = G$, where $X \sim F$, $Y \sim G$.

By a simple integral transform, a distortion measure ρ^g can be equivalently represented as

$$\rho^g(X) = \int_0^1 \text{V@R}_t(X)\,d\bar{g}(t),$$

where $\bar{g}(t) = 1 - g(1 - t)$ (see Dhaene et al. (2012)) and the *value-at-risk*

$$\text{V@R}_\alpha(X) = F^{-1}(\alpha) := \inf\{x \in \mathbb{R} | \, P(X \leq x) \geq \alpha\}, \text{ for } \alpha \in (0, 1). \tag{2}$$

We refer to $\bar{g}$ as a *flipped distortion function.*

The family of all distortion measures is convex and its extremals are given by the *average value-at-risk.*

Definition 2.3. The AV@R of a random variable X at confidence level $\alpha \in (0, 1)$ is defined as

$$\text{AV@R}_\alpha(X) := \frac{1}{1-\alpha} \int_\alpha^1 \text{V@R}_t(X)\,dt,$$

provided that the integral exists. It is the distortion risk measure pertaining to the functions

$$g_\alpha(t) = \min\left(\frac{t}{1-\alpha}, 1\right) \quad \text{and} \quad \bar{g}_\alpha(t) = \max\left(\frac{t-\alpha}{1-\alpha}, 0\right).^2 \tag{3}$$

To see that the extremal distortion functionals are AV@Rs, consider the *Kusuoka representation* (also called *Choquet representation*)

$$\rho^g(X) = \int_0^1 \text{AV@R}_\alpha(X)\,d\nu(\alpha),$$

where the relation between the probability measure ν on $[0, 1]$ and the flipped distortion function $\bar{g}$ is

$$\bar{g}(t) = 1 - g(1 - t) = \int_0^1 \frac{t - y}{1 - y}\,d\nu(y).$$

The proof is provided in Pflug and Pichler (2016), Chapter 3.

While the AV@R has the dual representation

$$\text{AV@R}_\alpha(X) = \sup\{\mathbb{E}(X \cdot Z) : 0 \leq Z \leq 1/(1 - \alpha),\, \mathbb{E}(Z) = 1\},$$

any distortion functional has the dual representation

$$\rho^g(X) = \sup\{\mathbb{E}(X \cdot Z) : (1 - \alpha)\text{AV@R}_\alpha(Z) \leq 1 - \bar{g}(\alpha),\, \mathbb{E}(Z) = 1\}.$$

(see Pflug and Pichler (2016), theorem 3.16). This representation as the maximum of linear functionals shows that ρ^g is convex in X; see proposition below.

[1] Two random variables X and Y are *comonotone* if they can be represented as $X = F^{-1}(U)$ and $Y = G^{-1}(U)$, respectively, with the same $U \sim \text{Uniform}[0, 1]$.

[2] The value-at-risk V@R is not a distortion functional in our sense, since it cannot be represented in form (1). Zhuang et al. (2016) also call the V@R a "distortion functional." Notice that there are also examples where $\text{V@R}_\alpha < \mathbb{E}(X)$, even for α arbitrarily close to 1.

ACCEPTED MANUSCRIPT

Proposition 2.1 (Pflug and Pichler (2016), theorem 3.27)**.** All distortion measures with concave distortion function g, and in particular the $\mathrm{AV@R}_\alpha$, enjoy the following properties:

1. ρ^g is convex in the random variable: $\rho^g(\lambda X + (1-\lambda)Y) \le \lambda\rho^g(X) + (1-\lambda)\rho^g(Y)$, for $0 \le \lambda \le 1$.
2. ρ^g is compound concave in the probability distribution: if $Y = X_1$ with probability λ and $Y = X_2$ with probability $1-\lambda$, then

$$\rho^g(Y) \ge \lambda\rho^g(X_1) + (1-\lambda)\rho^g(X_2), \quad \text{for } 0 \le \lambda \le 1.$$

As the distortion risk measure ρ^g depends on the underlying probability distribution F, the notion of robustness plays an important role when evaluating ρ^g under different distributions.

Definition 2.4. Let D be a distance for distribution functions. A distortion risk measure ρ^g is *robust (continuous)* w.r.t. the distance D if for $\forall X, Y \in L^1$, $X \sim F, Y \sim G$, $\forall \epsilon > 0$, there exists some $\delta > 0$ such that $D(F,G) \le \delta$ implies $\left|\rho^g(X) - \rho^g(Y)\right| \le \epsilon$.

Distortion risk premium. Distortion risk measures are also widely used as insurance premium principles; in fact, their origin lies in the premium calculation introduced by Denneberg (1990). The derivative $\overline{g}'$ of $\overline{g}$ is also called the *loading function*.

Definition 2.5. Let $g : [0,1] \to [0,1]$ be a distortion function. The *distortion premium* $\pi^{g,\theta}$ of the loss random variable X with distribution F is defined as

$$\pi^{g,\theta}(X) = (1+\theta)\int_0^\infty g(1-F(x))\,dx,$$

with constant $\theta \ge 0$ called *safety loading* of the insurer. Using the flipped distortion $\overline{g}(t) = 1 - g(1-t)$, the distortion premium principle can be equivalently written as

$$\pi^{g,\theta}(X) = (1+\theta)\int_0^1 \mathrm{V@R}_t(X)\,d\overline{g}(t).$$

Wang et al. (1997) proved that any premium principle that is equivariant, comonotone additive, positive homogeneous, and continuous in the following sense

$$\lim_{d\to 0}\pi(\max(X-d,0)) = \pi(X) \quad \text{and} \quad \lim_{d\to\infty}\pi(\min(X,d)) = \pi(X),$$

is a distortion premium. If g is concave, then $\pi^{g,\theta}(X) \ge \mathbb{E}(X)$, which on average ensures insurer survival.

Extreme value theory. The management of insurance companies relies on the necessity to precisely quantify the risk, namely, the probability of occurrence and the magnitude of the associated losses. The problem becomes crucial in the case of extreme events. Extreme value theory (EVT) represents the statistical framework needed to model low probability-high consequence events and to compute a measure for extreme risk.

Typically there are two ways of modeling extreme distributions:

- The *block maxima* approach considers the *sample maxima* $M_n = \max(X_1, X_2, \dots, X_n)$ for an i.i.d. sequence $(X_i)_{i\ge 1}$. By the Fisher-Tipett Theorem, extended by Gnedenko (Fisher and Tippett (1928), Gnedenko (1943)), the only non-degenerate limiting distributions H of the standardized sequence $c_n^{-1}(M_n - d_n)$ are of the form

$$H_\xi(x) = \begin{cases} \exp\{-(1+\xi x)^{-1/\xi}\}, & \text{if } \xi \ne 0, \\ \exp\{-\exp\{-x\}\}, & \text{if } \xi = 0, \end{cases}$$

  where $1 + \xi x > 0$ for shape parameter $\xi \in \mathbb{R}$. The one-parameter representation of H is known as *generalized extreme value distribution* (GEV). It includes the three types of extreme value distributions, i.e., Fréchet with $\xi > 0$ characteristic to heavy-tailed distributions, Gumbel with $\xi = 0$ for thin-tailed distributions, and Weibull with $\xi < 0$ for finite endpoint distributions.

- The *peak over threshold* (POT) approach studies the distribution of exceedances over a given threshold. By the Balkema-de Haan Theorem (Balkema and Haan (1974), Pickands et al. (1975)), the *excess distribution* $F_{\tilde{u}}(x) = P(X - \tilde{u} \le x | X > \tilde{u})$ satisfies

$$F_{\tilde{u}}(x) \to G,\ \tilde{u} \to \infty,$$

ACCEPTED MANUSCRIPT

where G is the *generalized Pareto distribution* (GPD) with location $\mu \in \mathbb{R}$, scale $\sigma \in (0, \infty)$ (depending on $\tilde{u}$) and shape $\xi \in \mathbb{R}$ given by

$$G(x) = \begin{cases} 1 - \left(1 + \xi \dfrac{x-\mu}{\sigma}\right)^{-1/\xi}, & \text{for } \xi \neq 0, \\ 1 - \exp\left(-\dfrac{x-\mu}{\sigma}\right), & \text{for } \xi = 0, \end{cases}$$

for $x \geq \mu$ when $\xi \geq 0$, and $\mu \leq x \leq \mu - \xi/\sigma$ when $\xi < 0$.

For a detailed discussion on EVT, see Coles et al. (2001) and Embrechts et al. (2013).

3 Problem formulation

In this section we formulate the problem of optimal design for insurance contracts from the policy holder's point of view. For simplicity, we consider only a single-stage problem where the contracting time is 0 and the observation period is $[0, T]$. Let $X \geq 0$ be a random variable representing the accumulated loss over the observation period, where X has estimated distribution $\hat{F}$. We assume that the insured entity is susceptible to high losses caused by extreme events. We refer to $\hat{F}$ as the *baseline/underlying distribution* of X.

To mitigate potential losses, the person affected by catastrophic events is seeking an insurance contract in exchange for a premium $\pi^{g,\theta}(I(X))$, with $\theta > 0$ being the safety loading. Here $I(X)$ is the payment function associated with loss X, written as the insurance contract. The most common types of insurance contract are the proportional contract, with payment $I(X) = cX$, for $0 < c \leq 1$, the stop-loss contact with $I(X) = \min(X, d)$ and the entrance-stop loss contract with $I(X) = \min(\max(X - d_1, 0), d_2 - d_1)$. The last is also known in literature by the name stop-loss insurance contract with an upper limit or as a one-layer insurance contract.

The retained loss that still needs to be covered by the insured entity is $X - I(X) + \pi^{g,\theta}(I(X))$. For the set of admissible contracts, we follow the same line as Cheung et al. (2012), Chi and Tan (2013), Lo (2017b) and consider the set of feasible contracts of the form

$$\mathcal{I} := \big\{ I : \mathbb{R}_+ \to \mathbb{R}_+ : \, I \text{ is non-decreasing}, 0 \leq I(x) \leq x, |I(x) - I(y)| \leq |x - y|, x, y \in \mathbb{R}_+ \big\}.$$

The aim is to find the optimal insurance contract design that minimizes the risk associated with the retained loss. More precisely,

$$\begin{aligned} &\inf_{I \in \mathcal{I}} \quad \rho^{g_1}\Big(X - I(X) + \pi^{g,\theta}\big(I(X)\big)\Big) \\ &\text{s.t.} \quad \pi^{g,\theta}\big(I(X)\big) \leq B, \end{aligned} \tag{P$_1$}$$

where $B > 0$ is a fixed budget. Distortion function g_1 is used by the insurance buyer to quantify the risk associated with the retained loss, while the insurer uses the distortion g to compute the premium. One can observe that (P$_1$) assumes full knowledge of the underlying distribution $\hat{F}$, i.e., the non-ambiguous case. Problem (P$_1$) will be extended to the ambiguous case later in this section.

Zhuang et al. (2016) solved (P$_1$) for general distortion functions, using a reformulation of the admissible set $\mathcal{I}$. By definition, any $I \in \mathcal{F}$ is absolute continuous; hence there exists $h : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$I(x) = \int_0^x h(z)\, dz. \tag{4}$$

The function h is called the *marginal indemnification function* (MIF)(see Assa (2015)). In reality, the insurance market includes coverage limitations (see Doherty et al. (2013), Cummins and Mahul (2004)) or encounters a moral hazard when facing large claims (see Balbás et al. (2015)). From the optimization point of view, this means that we restrict the codomain of function h. The set of feasible MIFs therefore becomes

$$\mathcal{H} := \big\{ h : \mathbb{R}_+ \to [0,1] : \, 0 \leq h \leq 1 \text{ a.e. and } h \text{ is Lebesgue measurable} \big\}. \tag{5}$$

Thus $\mathcal{I}$ is reformulated as

$$\mathcal{I} = \left\{ I : \mathbb{R}_+ \to \mathbb{R}_+ : \, I(x) = \int_0^x h(z)\, dz, \forall x \in \mathbb{R}_+, h \in \mathcal{H} \right\}.$$

ACCEPTED MANUSCRIPT

If I is of the form (4), then $\rho^g(I(X)) = \int_0^\infty g\big(1-\hat{F}(z)\big)h(z)\,dz$ for some distortion function g (see lemma 2.1. in Zhuang et al. (2016)).

Our next proposition generalizes this result for the ambiguous case, where more than one loss distribution is compatible with the observed data. To emphasize the use of alternative models when evaluating risk measures, we write X^F for a random variable, which has distribution F.

The distributionally robust formulation of (P_1) is

$$\begin{aligned} &\inf_{I\in\mathcal{I}}\sup_{F\in\mathcal{C}} \quad \rho^{g_1}\Big(X^F - I(X^F) + \pi^{g,\theta}\big(I(X^{\hat{F}})\big)\Big) \\ &\text{s.t.} \qquad\quad \pi^{g,\theta}\big(I(X^{\hat{F}})\big) \le B. \end{aligned} \tag{P$_2$}$$

The insurance buyer considers an optimal contract $I(X)$ which minimizes the risk measure ρ^{g_1} of the retained loss for the convex hull of alternative models $F_1, F_2, \dots, F_m$, i.e. $\mathcal{C} := conv(F_1, F_2, \dots, F_m)$. The insured person is inclined to buy the resulting contract $I(X)$ as long as the associated premium does not exceed the budget B. The premium is constructed based on a concave distortion function g and is computed w.r.t. the baseline distribution $\hat{F} \in \mathcal{C}$.

For the above problem, we impose the following condition:

Assumption 3.1. There exists some $K < \infty$ such that $\rho^{g_1}(X^F) \le K$, $\forall F \in \mathcal{C}$.

Considering the properness of the distortion risk measure we have $\rho^{g_1}(X^F) \ge \mathbb{E}(X^F)$; thus the assumption 3.1 implies the finiteness of the first moment for all $F \in \mathcal{C}$.

Proposition 3.1. Let $\mathcal{C} = conv(F_1, \dots, F_m)$ be the convex hull of a set of alternative models and let $\hat{F} \in \mathcal{C}$ be a baseline model based on which the insurance premium is computed. Let $g, g_1 : [0,1] \to [0,1]$ be two concave distortion functions used to construct the premium $\pi^{g,\theta}$ and the risk measure ρ^{g_1}, respectively. Then there exists some $F^* \in \mathcal{C}$ such that the distributionally robust optimization problem (P_2) has an optimal insurance contract $I^*(x) = \int_0^x h^*(z)\,dz$, where

$$h^*(z) = \begin{cases} 0, & \text{if } g_1(1-F^*(z)) - (1+\theta+\eta^*)g(1-\hat{F}(z)) < 0, \\ \kappa(z), & \text{if } g_1(1-F^*(z)) - (1+\theta+\eta^*)g(1-\hat{F}(z)) = 0, \\ 1, & \text{if } g_1(1-F^*(z)) - (1+\theta+\eta^*)g(1-\hat{F}(z)) > 0, \end{cases}$$

for some κ, a Lebesgue measurable function with $0 \le \kappa(z) \le 1$ and for some $\eta^* \ge 0$ satisfying

$$\pi^{g,\theta}\big(I^*(X^{\hat{F}})\big) = B.$$

Proof. Due to the comonotone additivity and translation equivariance properties of ρ^{g_1}, the problem (P_2) is equivalent to

$$\left\| \begin{aligned} &\inf_{h\in\mathcal{H}}\sup_{F\in\mathcal{C}} \quad \int_0^\infty g_1(1-F(z))dz - \int_0^\infty g_1(1-F(z))h(z)dz + (1+\theta)\int_0^\infty g(1-\hat{F}(z))h(z)dz \\ &\text{s.t.} \qquad\qquad \int_0^\infty g(1-\hat{F}(z))h(z)dz \le \overline{B}, \end{aligned} \right.$$

where $\overline{B} = B(1+\theta)^{-1}$. To prove the existence of a saddle point (see definition 6.3 in appendix), it is necessary to check whether (P_2) satisfies the conditions in the minimax theorem (see Sion's minimax theorem 6.1 in appendix). The set $\mathcal{H}_\pi := \mathcal{H} \bigcap \Big\{h \in \mathcal{H} : \int_0^\infty g(1-\hat{F}(z))h(z)dz \le \overline{B}\Big\}$ is non-empty as $h = 0$ belongs to the intersection. Moreover, since the constraint in (P_2) is linear in h, then $\mathcal{H}_\pi$ is also convex.

It can easily be seen that $\mathcal{C}$ is closed in the topology of weak convergence, as it is the convex hull of finitely many distributions. Let K be as in assumption 3.1. For any $\delta > 0$ define $K_\delta := K/\delta < \infty$. By Markov inequality, we then have that for all $F \in \mathcal{C}$,

$$F(K_\delta) \ge 1 - \frac{\mathbb{E}(X^F)}{K_\delta} \ge 1 - \delta,$$

which implies (uniform) tightness of $\mathcal{C}$. Because $\mathcal{C}$ is closed and (by Prokhorov (1956), in appendix) relatively compact in the topology of weak convergence, then $\mathcal{C}$ is weakly compact.

ACCEPTED MANUSCRIPT

The objective function in (P_2) is continuous in h and F, linear in h and concave in F, while $\mathcal{H}_\pi$ is a convex set and $\mathcal{C}$ is a convex and compact set. By Sion's minimax theorem (see 6.1 in appendix) there exists a saddle point, i.e., $\exists\, F^* \in \mathcal{C}$ such that

$$\begin{aligned}
&\inf_{h\in\mathcal{H}_\pi}\max_{F\in\mathcal{C}} \int_0^\infty g_1(1-F(z))dz - \int_0^\infty g_1(1-F(z))h(z)dz + (1+\theta)\int_0^\infty g(1-\hat{F}(z))h(z)dz \\
&= \inf_{h\in\mathcal{H}_\pi} \int_0^\infty g_1(1-F^*(z))dz - \int_0^\infty g_1(1-F^*(z))h(z)dz + (1+\theta)\int_0^\infty g(1-\hat{F}(z))h(z)dz \\
&= \int_0^\infty g_1(1-F^*(z))dz - \sup_{h\in\mathcal{H}_\pi}\Big[\int_0^\infty g_1(1-F^*(z))h(z)dz - (1+\theta)\int_0^\infty g(1-\hat{F}(z))h(z)dz\Big].
\end{aligned}$$

The inner optimization problem in the last equality can be equivalently written as:

$$\left\| \begin{aligned} &\sup_{h\in\mathcal{H}} \int_0^\infty g_1(1-F^*(z))h(z)dz - (1+\theta)\int_0^\infty g(1-\hat{F}(z))h(z)dz \\ &\text{s.t.} \int_0^\infty g(1-\hat{F}(z))h(z)dz \le \overline{B}. \end{aligned}\right. \tag{$\mathrm{P_{inner}}$}$$

As problem ($\mathrm{P_{inner}}$) is linear in $h \in \mathcal{H}$, the strong duality holds. For a dual variable $\eta \ge 0$, the dual inner problem is

$$\inf_{\eta\ge 0}\sup_{h\in\mathcal{H}} \mathcal{L}(h,\eta) = \inf_{\eta\ge0}\sup_{h\in\mathcal{H}} \int_0^\infty \Big[g_1(1-F^*(z)) - (1+\theta+\eta)g(1-\hat{F}(z))\Big]h(z)\,dz + \eta\overline{B}. \tag{$\mathrm{D_{inner}}$}$$

Similar to Zhuang et al. (2016) in the case of a single distribution, define the sets:

$$\begin{cases} A^+ := \{z : g_1(1-F^*(z)) - (1+\theta+\eta)g(1-\hat{F}(z)) > 0\}, \\ A^0 := \{z : g_1(1-F^*(z)) - (1+\theta+\eta)g(1-\hat{F}(z)) = 0\}, \\ A^- := \{z : g_1(1-F^*(z)) - (1+\theta+\eta)g(1-\hat{F}(z)) < 0\}. \end{cases}$$

Define the MIF $h^* \in \mathcal{H}$, which depends on the value of $\eta \ge 0$, to be of the following form:

$$h^*(\cdot;\eta) = \begin{cases} 0, & \text{if } z \in A^-, \\ \kappa(z), & \text{if } z \in A^0, \\ 1, & \text{if } z \in A^+, \end{cases} \tag{6}$$

for some arbitrary $\kappa : [0,\infty) \to [0,1]$ Lebesgue measurable function.

The constraint in ($\mathrm{P_{inner}}$) only considers the baseline distribution; hence the existence of the dual variable $\eta^* \ge 0$ such that

$$\int_0^\infty g(1-\hat{F}(z))h^*(z;\eta^*)dz = \overline{B} \tag{7}$$

is guaranteed by theorem (3.1.) in Zhuang et al. (2016).

Then $\eta^* \ge 0$ and the corresponding $h^*(\cdot;\eta^*) \in \mathcal{H}$ of the form (6) are feasible for ($\mathrm{D_{inner}}$) and ($\mathrm{P_{inner}}$), respectively. Moreover, by construction, η^* and $h^*(\cdot;\eta^*)$ satisfy (7); hence by complementary slackness condition, η^* and $h^*(\cdot;\eta^*)$ are optimal solutions of ($\mathrm{D_{inner}}$) and ($\mathrm{P_{inner}}$), respectively.

The original problem (P_2) has an optimal value:

$$\int_0^\infty g_1(1-F^*(z))dz - \int_0^\infty g_1(1-F^*(z))h^*(z;\eta^*)dz + (1+\theta)\int_0^\infty g(1-\hat{F}(z))h(z;\eta^*)dz,$$

where the corresponding MIF h^* is of the form (6). □

If the risk measure ρ^{g_1} is AV@R$_\alpha$, for some $\alpha \in (0,1)$ and $F_1,\dots,F_m$ are continuous, strictly increasing distribution functions, then (P_2) has an explicit solution, as is shown below.

Proposition 3.2. Let $\mathcal{C} = conv(F_1, F_2, \dots, F_m)$ be the convex hull of a set of strictly increasing, continuous cumulative distribution functions and let $\hat{F} \in \mathcal{C}$ be a baseline distribution. Let $g : [0,1] \to [0,1]$ be a strictly increasing, concave distortion function. Then the optimal $I \in \mathcal{I}$ that solves problem

$$\begin{aligned} &\inf_{I\in\mathcal{I}}\sup_{F\in\mathcal{C}} \quad \mathrm{AV@R}_\alpha\Big(X^F - I(X^F) + \pi^{g,\theta}\big(I(X^{\hat{F}})\big)\Big) \\ &\text{s.t.} \qquad \pi^{g,\theta}\big(I(X^{\hat{F}})\big) \le B, \end{aligned} \tag{P_3}$$

ACCEPTED MANUSCRIPT

for some $\alpha \in (0,1)$ is an entrance excess-of-loss contract, i.e., there exist $d_1, d_2 \in \mathbb{R}_+$, $d_1 \le d_2$ such that

$$I^*(x) = \begin{cases} 0, & \text{if } 0 \le x \le d_1, \\ x - d_1, & \text{if } d_1 < x \le d_2, \\ d_2 - d_1, & \text{if } d_2 < x. \end{cases} \tag{8}$$

Proof. Using the minimax property proven in proposition 3.1, there exists some optimal $F^* \in \mathcal{C}$ such that the problem (P_3) can be reformulated as follows:

$$\left\| \begin{array}{l} \displaystyle\int_0^\infty g_\alpha(1 - F^*(z))\, dz - \sup_{h\in\mathcal{H}} \Big[\int_0^\infty \big[g_\alpha\big(1 - F^*(z)\big) - (1+\theta) g\big(1 - \hat{F}(z)\big)\big] h(z)\, dz \Big] \\ \text{s.t.} \displaystyle\int_0^\infty g\big(1 - \hat{F}(z)\big)\, h(z)\, dz \le \overline{B}, \end{array} \right.$$

where $\overline{B} = (1+\theta)^{-1} B$ and $g_\alpha(z) = \min\left(\frac{z}{1-\alpha}, 1\right)$ is the distortion function corresponding to AV@R$_\alpha$. Again, the inner problem is

$$\left\| \begin{array}{l} \displaystyle\sup_{h\in\mathcal{H}} \int_0^\infty \big[g_\alpha\big(1 - F^*(z)\big) - (1+\theta) g\big(1 - \hat{F}(z)\big)\big] h(z)\, dz \\ \text{s.t.} \displaystyle\int_0^\infty g\big(1 - \hat{F}(z)\big)\, h(z)\, dz \le \overline{B}. \end{array} \right. \tag{P'_{inner}}$$

If $\hat{h} := \mathbb{1}\{g_\alpha\big(1 - F^*(z)\big) - (1+\theta) g\big(1 - \hat{F}(z)\big) > 0\}$ satisfies the constraint in (P'_{inner}), then it is the optimal solution of (P'_{inner}). Otherwise, due to linearity in h, the strong duality holds with the dual problem

$$\inf_{\eta \ge 0} \sup_{h\in\mathcal{H}} \int_0^\infty g(1 - \hat{F}(z)) \left(\frac{g_\alpha(1 - F^*(z))}{g(1 - \hat{F}(z))} - (1+\theta+\eta) \right) h(z)\, dz + \eta \overline{B}.$$

Denote $\mathcal{G}(z) := \dfrac{g_\alpha\big(1 - F^*(z)\big)}{g\big(1 - \hat{F}(z)\big)} = \begin{cases} \dfrac{1}{g\big(1 - \hat{F}(z)\big)}, & \text{if } 0 \le z \le \text{V@R}_\alpha(X^{F^*}), \\ \dfrac{1 - F^*(z)}{(1-\alpha) g\big(1 - \hat{F}(z)\big)}, & \text{if } \text{V@R}_\alpha(X^{F^*}) < z. \end{cases}$

Since F^* is continuous and g is a concave function, then $\mathcal{G}$ is continuous, increasing on $[0, \text{V@R}_\alpha(X^{F^*})]$ and decreasing on $(\text{V@R}_\alpha(X^{F^*}), \infty)$ with $\mathcal{G}(\text{V@R}_\alpha(X^{F^*})) = g\big(1 - \hat{F}(\text{V@R}_\alpha(X^{F^*}))\big)^{-1} \ge 1$. Moreover, $\lim_{z\to 0} \mathcal{G}(z) = 1$ and $\lim_{z\to\infty} \mathcal{G}(z) = 0$. Since $g\big(1 - \hat{F}(z)\big) \ge 0$, to determine the optimal $h \in \mathcal{H}$, one need only study the sign of $\mathcal{G}(z) - (1+\theta+\eta)$. We distinguish two cases.

Case 1. If $1 + \theta + \eta < \mathcal{G}\big(\text{V@R}_\alpha(X^{F^*})\big)$, then by the intermediate value theorem, there exists some $d_1 \in [0, \text{V@R}_\alpha(X^{F^*})]$ and $d_2 \in (\text{V@R}_\alpha(X^{F^*}), \infty)$, such that $\mathcal{G}(d_1) = \mathcal{G}(d_2) = 1 + \theta + \eta$. Then the optimal h^*, which depends on η, has the following form:

$$h^*(z;\eta) = \begin{cases} 0, & \text{if } z < d_1, \\ 1, & \text{if } d_1 \le z < d_2, \\ 0, & \text{if } d_2 \le z. \end{cases} \tag{9}$$

This defines $I(x) = \int_0^x h^*(z;\eta) dz$ as in (8).

Case 2. If $1 + \theta + \eta \ge \mathcal{G}\big(\text{V@R}_\alpha(X^{F^*})\big)$, then $h^* = 0$. In this case, $d_1 = d_2 = \infty$. We are going to prove later on that this case is not possible, when taking into account the constraint in (P'_{inner}).

The existence of $\eta \ge 0$ such that $\int_0^\infty g\big(1 - \hat{F}(z)\big) h^*(z;\eta) = \overline{B}$ is proven in theorem (4.1.) in Zhuang et al. (2016). This implies that $h^* \ne 0$. Again, (P_3) has an optimal value given by

$$\int_0^\infty g_\alpha(1 - F^*(z))\, dz - \int_0^\infty \big[g_\alpha\big(1 - F^*(z)\big) - (1+\theta) g\big(1 - \hat{F}(z)\big)\big] h^*(z)\, dz.$$

□

Figure (1) illustrates the payment function of the contract $I(X) = \min(\max(X - d_1, 0), d_2 - d_1)$. The entrance excess-of-loss insurance contracts are proven to be optimal in the contexts of very large claims and the coverage limitations of the insurance market (see Cummins and Mahul (2004) and Doherty et al. (2013)).

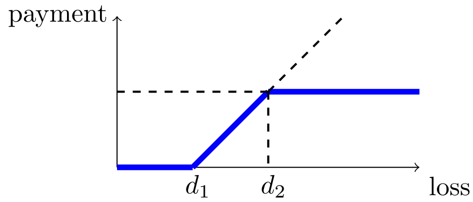


Figure 1: Insurance contract with deductible d_1 and cap d_2.

4 Alternative models

In the distributionally robust problems (P_2) and (P_3), the set of alternative models is given a priori, without any further specification. In this section, we will discuss a method of generating the alternative models and of finding the optimal parameters of the contract obtained in proposition 3.2.

For $r \geq 1$, let F and G be two distributions on $(\Omega, \mathcal{F}, P)$ with finite moments of order r.

Definition 4.1. The *Wasserstein distance* of order r between probability distributions F and G is

$$\mathrm{WD}_{d,r}(F,G) := \inf_{\substack{X \sim F \\ Y \sim G}} \left[\mathbb{E}\, d(X,Y)^r\right]^{1/r},$$

where the infimum is among all joint probabilities on $\Omega \times \Omega$ with fixed marginals F and G. Here d is a metric on $\mathbb{R}$. Typically d is the 1-norm, i.e., $d(x,y) = |x - y|$.

The Wasserstein distance satisfies the triangle inequality and enjoys the following properties:

- If $r_1 \leq r_2$, then $\mathrm{WD}_{d,r_1}(F,G) \leq \mathrm{WD}_{d,r_2}(F,G)$.
- $\mathrm{WD}_{d,r}$ is symmetric and convex in both arguments, i.e., for $0 \leq \lambda \leq 1$,

$$\mathrm{WD}_{d,r}(F, \lambda G_1 + (1-\lambda) G_2)^r \leq \lambda \mathrm{WD}_{d,r}(F, G_1)^r + (1-\lambda)\mathrm{WD}_{d,r}(F, G_2)^r.$$

For more on interpretation and properties of Wasserstein distance, see Villani (2008), Chapter 6.

The Wasserstein distance of order $r \geq 1$ in the case $\Omega = \mathbb{R}$ with 1-norm is given by

$$\mathrm{WD}_{1,r}(F,G) = \left[\int_{-\infty}^{\infty} \left|F(x) - G(x)\right|^r dx\right]^{1/r} = \left[\int_0^1 \left|F^{-1}(y) - G^{-1}(y)\right|^r dy\right]^{1/r}. \tag{10}$$

For a proof for $r = 1$, see Vallender (1974). The general case $r \geq 1$ can be proven in a similar way.

The average value-at-risk is robust with respect to 1-Wasserstein distance in the sense of definition 2.4:

$$|\mathrm{AV@R}_\alpha(X^F) - \mathrm{AV@R}_\alpha(X^G)| \leq \frac{1}{1-\alpha}\mathrm{WD}_{1,1}(F,G). \tag{11}$$

See Kiesel et al. (2016).

From (10) we can observe that $\mathrm{WD}_{1,1}$ assigns equal weight to the difference between the distributions F and G. The idea of replacing the Euclidean distance on $\mathbb{R}$ with another distance is motivated by observation of insurance for extreme events. More precisely, when we consider the order statistics of losses $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$, the difference between low losses (e.g., the distance between the first and second smallest observations) should not be seen as equal to the difference between very high losses (e.g., the distance between the 100-th and 101-th largest observations) from the insurance pricing point of view. The reason is that in extreme events, we would impose a higher penalty on the deviations from the baseline model at high quantiles while allowing differences between models around the mean of the distribution. We therefore need to define a metric which is more sensitive to the tail of the distribution.

For this reason, we propose the following transformation of the positive real line:

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Definition 4.2. Let $x_q \in \mathbb{R}_+$ fixed and let $\varphi_{s,x_q} : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{\geq 0}$ be a bijective transformation of the positive real line defined by

$$\varphi_{s,x_q}(x) = \begin{cases} x, & \text{if } x \leq x_q \\ x_q^{1-s} x^s, & \text{else} \end{cases}, \quad s \in \mathbb{N},$$

which induces the metric $d_{s,x_q}(x,y) := |\varphi_{s,x_q}(x) - \varphi_{s,x_q}(y)|$.

Figure (2) indicates the manner in which the transformation φ_{s,x_q} contorts $[0,\infty)$: the values smaller than the constant x_q are unchanged, while the larger values are inflated. This property turns out to be appropriate for extreme value analysis, where the focus is on the shape parameter ξ.

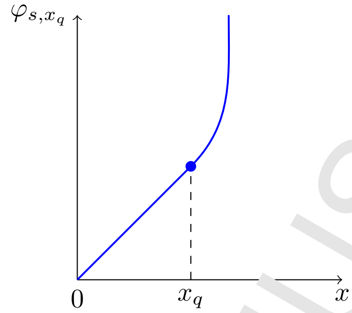


Figure 2: Transformation φ_{s,x_q}.

Remark 4.1.

1. In extreme value analysis, q is a high probability from which we consider that the tail of all the models begins and $x_q := \hat{F}^{-1}(q)$. We can assume without loss of generality that $x_q >> 1$.
2. For $s \geq 2$, d_{s,x_q} and $d(x,y) = |x-y|$ are equivalent on bounded intervals. For a proof, see appendix.

If $X \sim F$ is a random variable with support $[0,\infty)$, then $\varphi_{s,x_q}(X) \sim F_{s,x_q}$ with

$$F_{s,x_q}(x) = P(\varphi_{s,x_q}(X) \leq x) = \begin{cases} F(x), & \text{if } X(\omega) < x_q, \\ F(x^{1/s} x_q^{1-1/s}), & \text{else} \end{cases}, s \in \mathbb{N}.$$

Definition 4.3. The Wasserstein distance of order 1 with the underlying metric given by d_{s,x_q} between the probability measures F and G with finite first s moments, is defined as

$$\mathrm{WD}_{d_{s,x_q},1}(F,G) = \inf_{\substack{X \sim F \\ Y \sim G}} \mathbb{E}[d_{s,x_q}(X,Y)] \tag{12}$$

and is called *contorted Wasserstein distance* between F and G.

Remark 4.2. Using a similar proof as in Vallender (1974), the contorted Wasserstein distance between probability distributions F and G has the following form

$$\begin{aligned} \mathrm{WD}_{d_{s,x_q},1}(F,G) &= \int_0^\infty \big|F(x) - G(x)\big| \varphi'_{s,x_q}\, dx \\ &= \int_0^{x_q} \big|F(x) - G(x)\big|\, dx + \int_{x_q}^\infty \big|F(x) - G(x)\big| s(x/x_q)^{s-1}\, dx. \end{aligned} \tag{13}$$

From now on, as all the distances considered are of order 1, we omit the order in the notation.

Remark 4.3. If $x_q \geq 1$, $|\varphi'_{s,x_q}| \geq 1$, then $\mathrm{WD}_1(F,G) \leq \mathrm{WD}_{d_{s,x_q}}(F,G)$, for any probability distributions F and G. Moreover, the contorted Wasserstein distance satisfies the same properties as WD_1.

Proposition 4.1. The contorted Wasserstein distance satisfies the following properties:

1. $\mathrm{WD}_{d_{s,x_q}}$ characterizes the weak topology on sets of distributions with uniformly bounded s moments: let $(F_n)_{n\geq 1}$ be a sequence of distribution functions and F another distribution function. If F_n, F have bounded s moments, then

$$\mathrm{WD}_{d_{s,x_q}}(F_n, F) \xrightarrow[n\to\infty]{} 0 \quad \iff \quad F_n \xrightarrow[n\to\infty]{} F \text{ weakly.}$$

ACCEPTED MANUSCRIPT

2. Assume that the right endpoint of a probability distribution F is finite, i.e., ess $\sup(F) < \infty$. Then there exists some constant K such that
$$P\{\mathrm{WD}_{d_{s,x_q}}(\hat{F}_n, F) \geq \epsilon\} \leq K\epsilon^{-1}n^{-1},$$
where $\hat{F}_n(x) = \dfrac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(-\infty,x]}(X_{(i)})$ is the empirical distribution function on $\mathbb{R}$ of the i.i.d. sample $\{X_{(1)}, X_{(2)}, \ldots, X_{(n)}\}$ from a probability distribution F.

3. Let $\big(X_i, Y_i\big)_{i\in\mathbb{N}}$ and $\big(\widetilde{X}_i, Y_i\big)_{i\in\mathbb{N}}$ be two renewal models (as in definition 6.1 in appendix) with the same claim times Y_i. Let F and G be the distribution functions of X_i and $\tilde{X}_i$, respectively. If $\overline{F} = 1 - F$ and $\overline{G} = 1 - G$ are regularly varying functions, then for large enough initial capital u, the ruin probability $\psi(u, \cdot)$ (see definition 6.2 in appendix) satisfies
$$|\psi(u, F) - \psi(u, G)| \leq C \cdot \mathrm{WD}_{d_{s,x_q}}(F, G),$$
for some positive constant C.

Proof. 1. Since d_{s,x_q} is a distance on $\mathbb{R}_+$, the property follows in a similar way as in Villani (2008), theorem 7.2.

2. If $M :=$ ess $\sup(F)$, then $\int_M^\infty |\hat{F}_n(x) - F(x)|\, dx = 0$. The distorted Wasserstein distance is
$$\begin{aligned}\mathrm{WD}_{d_{s,x_q}}(\hat{F}_n, F) &= \int_0^M \big|\hat{F}_n\big(\varphi^{-1}_{s,x_q}(x)\big) - F\big(\varphi^{-1}_{s,x_q}(x)\big)\big|\, dx = \int_0^{\varphi^{-1}_{s,x_q}(M)} \big|\hat{F}_n(x) - F(x)\big|\varphi'_{s,x_q}(x)\, dx \\ &\leq \varphi'_{s,x_q}(M) \int_0^{\varphi^{-1}_{s,x_q}(M)} \big|\hat{F}_n(x) - F(x)\big|\, dx \\ &= \varphi'_{s,x_q}(M)\, \mathrm{WD}_1(\hat{F}_n, F) < \infty.\end{aligned}$$
$\mathbb{E}\big[\mathrm{WD}_{d_{s,x_q}}(\hat{F}_n, F)\big] \leq \varphi'_{s,x_q}(M)\mathbb{E}\big[\mathrm{WD}_1(\hat{F}_n, F)\big] \leq C \cdot n^{-1}$, for some constant C, where the last inequality holds by Dudley (1969). Applying the Markov inequality yields the desired result.

3. The proof is straightforward and is presented in the appendix.

□

From now on, we fix $q \in (0, 1)$ and $s \in \mathbb{N}$. Since $x_q = \hat{F}^{-1}(q)$ is related only to the baseline distribution $\hat{F}$, we may omit the subscript x_q and, for the sake of simplicity, write $\mathrm{WD}_{d_{s,q}}$. The corresponding ambiguity set around $\hat{F}$ of radius $\epsilon > 0$ will be specified by the Wasserstein ball $\mathcal{P}_\epsilon(\hat{F}) := \{F : \mathrm{WD}_{d_{s,q}}(F, \hat{F}) \leq \epsilon\}$.

Remark 4.4. If the alternative models $F_1, F_2, \ldots, F_m \in \mathcal{P}_\epsilon(\hat{F})$, then the compactness of the set $\mathcal{C} = conv(F_1, \ldots, F_m)$ in proposition 3.1 and 3.2 is guaranteed by the compactness of $\mathcal{P}_\epsilon(\hat{F})$. To see this, observe that $\mathcal{C}$ is a subset of $\mathcal{P}_\epsilon(\hat{F})$, since any $F \in \mathcal{C}$ can be written as $F = \sum_{i=1}^m \lambda_i F_i$, for some $\lambda_i \geq 0$ with $\sum_{i=1}^m \lambda_i = 1$ and hence
$$\mathrm{WD}_{d_{s,q}}(F, \hat{F}) = \mathrm{WD}_{d_{s,q}}\Big(\sum_{i=1}^m \lambda_i F_i, \hat{F}\Big) \leq \sum_{i=1}^m \lambda_i \mathrm{WD}_{d_{s,q}}(F_i, \hat{F}) \leq \epsilon.$$
We obtain that $\mathcal{C}$ is a closed subset of a compact set $\mathcal{P}_\epsilon(\hat{F})$, and hence compact.

Furthermore, if the risk measure ρ^{g_1} is robust w.r.t. $\mathrm{WD}_{d_{s,q}}$ (as in definition 2.4), then for any $F \in \mathcal{C}$ and any $\delta > 0$, $\mathrm{WD}_{d_{s,q}}(F, \hat{F}) < \epsilon$ implies $\rho^{g_1}(X^F) < \rho^{g_1}(X^{\hat{F}}) + \delta$. Therefore, the condition that $\rho^{g_1}(X^F) < \infty$, for all $F \in \mathcal{C}$ reduces to the assumption that $\rho^{g_1}(X^{\hat{F}})$ is finite only under the baseline model.

Problem (P3) can be further extended by replacing the feasible set $\mathcal{C}$ by $\mathcal{P}_\epsilon(\hat{F})$.

Proposition 4.2. Let $\hat{F}$ be a baseline distribution and $g, g_1 : [0, 1] \to [0, 1]$ be some concave distortion functions. Then there exists some $F^* \in \mathcal{P}_\epsilon(\hat{F})$ and $h \in \mathcal{H}$ such that the optimal insurance contract $I^* \in \mathcal{I}$ of the following problem
$$\begin{aligned}&\inf_{I\in\mathcal{I}} \sup_{F\in\mathcal{P}_\epsilon(\hat{F})} \quad \rho^{g_1}\Big(X^F - I(X^F) + \pi^{g,\theta}\big(I(X^{\hat{F}})\big)\Big) \\ &\text{s.t.} \qquad \pi^{g,\theta}\big(I(X^{\hat{F}})\big) \leq B\end{aligned} \tag{P$_4$}$$
has a MIF h^* of the form (6).

ACCEPTED MANUSCRIPT

Proof. For $h \in \mathcal{H}_\pi = \mathcal{H} \bigcap \Big\{ h \in \mathcal{H} : \int_0^\infty g(1-\hat{F}(z))h(z)dz \leq \overline{B} \Big\}$ and a probability distribution F, let

$$\Gamma(h,F) := \rho^{g_1}\big(X^F - I(X^F) + \pi^{g,\theta}\big(I(X^{\hat{F}})\big)\big),$$

where $I(\cdot)$ is given by (4). Note that both $\mathcal{P}_\epsilon(\hat{F})$ and $\mathcal{H}_\pi$ are convex sets, $\mathcal{P}_\epsilon(\hat{F})$ is compact w.r.t. the contorted Wasserstein distance and the function $\Gamma(h,F)$ is linear in h and concave in F, therefore applying again Sion et al. (1958) yields:

$$\inf_{h\in\mathcal{H}_\pi} \max_{F\in\mathcal{P}_\epsilon(\hat{F})} \Gamma(h,F) = \max_{F\in\mathcal{P}_\epsilon(\hat{F})} \inf_{h\in\mathcal{H}_\pi} \Gamma(h,F).$$

The structure of the optimal $h^* \in \mathcal{H}$ is then proven in a similar way to proposition 3.1. □

Denote by $\mathcal{C}(\epsilon)$ the convex hull of some distributions $F_1,\ldots,F_m \in \mathcal{P}_\epsilon(\hat{F})$. The next result gives a bound on the optimal value of (P_2) when increasing the ambiguity radius ϵ.

Proposition 4.3. Let $g_1 : [0,1] \to [0,1]$ be a concave distortion function such that the flipped distortion function $\overline{g}_1(z) = 1 - g_1(1-z)$ satisfies $\|\overline{g_1}'\|_\infty < \infty$. For $0 < \epsilon_2 < \epsilon_1$ ambiguity radii, then there exists some $\delta > 0$ such that

$$\Big|\min_{h\in\mathcal{H}} \max_{F\in\mathcal{C}(\epsilon_1)} \Gamma(h,F) - \min_{h\in\mathcal{H}} \max_{F\in\mathcal{C}(\epsilon_2)} \Gamma(h,F)\Big| \leq 2(\epsilon_1 + \epsilon_2)\|\overline{g_1}'\|_\infty.$$

Proof. For $h \in \mathcal{H}$, define $\Gamma_{\epsilon_1}(h) := \max_{F\in\mathcal{C}(\epsilon_1)} \Gamma(h,F)$ and $\Gamma_{\epsilon_2}(h) := \max_{F\in\mathcal{C}(\epsilon_2)} \Gamma(h,F)$. Since $\Gamma_{\epsilon_i}(h)$ is concave in F, one can find $F_i^* = \operatorname{argmax}\{\Gamma_{\epsilon_i}(h) : F \in \mathcal{C}(\epsilon_i)\}$, for $i = 1,2$ and for given $h \in \mathcal{H}$. Because $F_i^* \in \mathcal{P}_{\epsilon_i}(\hat{F})$, then by triangle inequality of the contorted Wasserstein distance,

$$\mathrm{WD}_{d_{s,q}}(F_1^*, F_2^*) \leq \mathrm{WD}_{d_{s,q}}(F_1^*, \hat{F}) + \mathrm{WD}_{d_{s,q}}(F_2^*, \hat{F}) = \epsilon_1 + \epsilon_2.$$

Then the following holds

$$\begin{aligned}
\big|\Gamma_{\epsilon_1}(h) - \Gamma_{\epsilon_2}(h)\big| &= \big|\max_{F\in\mathcal{C}(\epsilon_1)} \Gamma(h,F) - \max_{F\in\mathcal{C}(\epsilon_2)} \Gamma(h,F)\big| \\
&= \big|\Gamma(h,F_1^*) - \Gamma(h,F_2^*)\big| \\
&\leq \big|\rho^{g_1}(X^{F_1^*}) - \rho^{g_1}(X^{F_2^*})\big| + \big|\rho^{g_1}(I(X^{F_1^*})) - \rho^{g_1}(I(X^{F_2^*}))\big| \\
&\leq \int_0^\infty \big|g_1(1-F_1^*(z)) - g_1(1-F_2^*(z))\big|dz + \int_0^\infty \big|g_1(1-F_2^*(z)) - g_1(1-F_2^*(z))\big|h(z)dz \\
&\leq (1+\|h\|_\infty)\int_0^\infty \big|g_1(1-F_1^*(z)) - g_1(1-F_2^*(z))\big|dz \\
&= (1+\|h\|_\infty)\int_0^1 \big|\big((F_1^*)^{-1}(z) - (F_2^*)^{-1}(z)\big)\overline{g}_1(z)\big|dz \\
&\leq (1+\|h\|_\infty)\|\overline{g}_1\|_\infty \int_0^1 \big|(F_1^*)^{-1}(z) - (F_2^*)^{-1}(z)\big|dz \\
&\leq (1+\|h\|_\infty)\|\overline{g}_1'\|_\infty \mathrm{WD}_1(F_1^*, F_2^*) \\
&\leq (1+\|h\|_\infty)\|\overline{g}_1'\|_\infty(\epsilon_1+\epsilon_2).
\end{aligned}$$

Without loss of generality, assume that $\epsilon_2 < \epsilon_1$. Then choose $\tilde{h} \in \mathcal{H}$ such that $\Gamma_{\epsilon_2}(\tilde{h}) \leq \min_{h\in\mathcal{H}} \Gamma_{\epsilon_2}(\tilde{h}) + \widetilde{\epsilon}$, for some $\widetilde{\epsilon} > 0$. Then

$$\begin{aligned}
\min_{h\in\mathcal{H}} \Gamma_{\epsilon_1}(h) - \min_{h\in\mathcal{H}} \Gamma_{\epsilon_2}(h) &\leq \min_{h\in\mathcal{H}} \Gamma_{\epsilon_1}(h) - \Gamma_{\epsilon_2}(\tilde{h}) + \widetilde{\epsilon} \\
&\leq \Gamma_{\epsilon_1}(\tilde{h}) - \Gamma_{\epsilon_2}(\tilde{h}) + \widetilde{\epsilon} \\
&\leq (1+\|h\|_\infty)\|\overline{g}_1'\|_\infty(\epsilon_1+\epsilon_2) + \widetilde{\epsilon}.
\end{aligned}$$

Since $\widetilde{\epsilon}$ is arbitrary and $\max_{h\in\mathcal{H}} \|\tilde{h}\|_\infty = 1$, the result follows. □

ACCEPTED MANUSCRIPT

Minimax algorithm. For the numerical section, we assume the continuity of the distribution functions. The convex hull of the set of alternative models $\mathcal{C}$ is constructed in a dynamic way via the following distributionally robust optimization problem:

$$\begin{aligned} \min_{h\in\mathcal{H}} \max_{F} \quad & \mathrm{AV@R}_\alpha(X^F - I(X^F) + \pi^{g,\theta}(I(X^{\hat{F}}))) \\ \text{s.t.} \quad & \pi^{g,\theta}(I(X^{\hat{F}})) \le B \\ & \mathrm{WD}_{d_{s,q}}(F, \hat{F}) \le \epsilon. \end{aligned} \tag{P$_5$}$$

Since the risk measure in the objective function of the problem (P$_5$) depends on alternative probability distribution F, as well as on the MIF h, the minimax problem is solved in a stepwise manner (see Pflug and Wozabal (2007)). The inner problem is of the form:

$$\begin{aligned} \max_{F} \ & \mathrm{AV@R}_\alpha(X^F - I(X^F) + \pi^{g,\theta}(I(X^{\hat{F}}))) \\ \text{s.t.} \ & \mathrm{WD}_{d_{s,q}}(F, \hat{F}) \le \epsilon \end{aligned}$$

and requires as input some $h \in \mathcal{H}$. It is a convex optimization problem (see proposition 2.1); hence there exists some F^* that maximizes $\mathrm{AV@R}_\alpha(\cdot)$. The outer problem

$$\begin{aligned} \min_{h\in\mathcal{H}} \max_{F\in\mathcal{C}} \ & \mathrm{AV@R}_\alpha(X^F - I(X^F) + \pi^{g,\theta}(I(X^{\hat{F}}))) \\ \text{s.t.} \ & \pi^{g,\theta}(I(X^{\hat{F}})) \le B \end{aligned}$$

is linear in h. The minimax procedure is then the following: in the initialization step, the admissible set $\mathcal{C}$ contains only the baseline distribution $\hat{F}$; hence, the outer problem is a non-ambiguous problem for which the optimal MIF is of the form $h(z) = \mathbb{1}\{(d_1^{(0)}, d_2^{(0)}]\}$ for some values $d_1^{(0)} \le d_2^{(0)}$. The inner problem is solved with parameters $d_1^{(0)}$ and $d_2^{(0)}$ as input, and the worst-case model F_1 is computed via convex optimization-based algorithm. The new-found model F_1 is added to $\mathcal{C}$ and the outer problem is solved, where the maximum is chosen w.r.t. the enlarged admissible set, i.e., $\mathcal{C} = conv(\hat{F}, F_1)$. The optimal insurance contract over $\mathcal{C}$ in the outer problem is of the form (8) for some $d_1^{(1)} \le d_2^{(1)}$, according to proposition 3.2. The optimal solution $(d_1^{(1)}, d_2^{(1)})$ at this iteration will be used again as input for the next inner problem. The procedure stops when the number of alternative models reaches m.

One of the difficulties in the minimax problem lies in the computation of the contorted Wasserstein distance between the baseline distribution $\hat{F}$ and the alternative distribution F. In spite of the compact form of this distance (see (13)), it is impossible to determine the integral in an analytic way, unless strong assumptions regarding the class of distributions considered are imposed. For instance, if all the probability distributions are discrete, then a linear programming approximation to compute $\mathrm{WD}_{d_{s,q}}$ can be formulated. The quality of the approximation depends on the discretization technique, i.e., optimal trade-off between a finer discretization and the numerical challenges faced when evaluating it. The problem can become even more difficult when discretizing the tail of the distributions, where the extreme events lay. To tackle these problems, we propose the following representation for the models in the ambiguity set.

Assumption 4.1. The baseline model $\hat{F}$ is piece-wise linear until x_q with a finite number of breakpoints $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$, where $x^{(n)} = x_q$. For $x \ge x_q$, we assume that $\hat{F}$ has a Pareto type tail, i.e., $1 - \hat{F}(x) = \hat{c}x^{-1/\hat{\xi}}$, for a constant $\hat{c} > 0$ and shape parameter $0 \le \hat{\xi} \le 1$. These values are known a priori by estimating the baseline distribution. An alternative distribution $F \in \mathcal{P}_\epsilon(\hat{F})$ is assumed to have a similar structure, i.e., F is piece-wise linear between $(x^{(i)}, x^{(i+1)})$, $i = \overline{0,n}$, but has different probabilities $0 = F^{(0)} < F^{(1)} < F^{(2)} < \ldots < F^{(n-1)}$. As the tail of the alternative distribution is assumed to start from the same x_q, we required that $F^{(n)} = \hat{F}^{(n)}$ and that it is of Pareto type, i.e., $1 - F(x) = cx^{-1/\xi}$, for some $c > 0$ and $\xi \in (0, 1)$.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

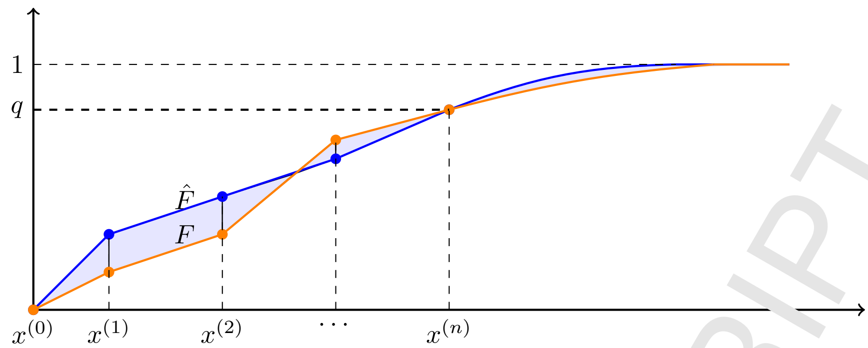


Figure 3: Distributions $\hat{F}$ and F satisfying assumption 4.1.

The assumption of Pareto tails for distributions in $\mathcal{P}_\epsilon(\hat{F})$ comes from EVT modeling of insurance losses. Since the extreme losses are considered positive random variables, their distribution is usually a heavy tailed one in the domain of attraction of the Fréchet distribution, with a shape parameter $\xi > 0$. Since $|\mathbb{E}(X^F) - \mathbb{E}(X^{\hat{F}})| \le \mathrm{WD}_{d_{s,q}}(F, \hat{F}) < \infty$, then the finiteness of the first moment implies $\xi < 1$. For a discussion on the typical values of parameter ξ in the insurance context, see Embrechts et al. (2013), Chapter 6.

In any ϵ-neighborhood w.r.t. the WD_1 of some baseline distribution with Pareto-like tail, one can find distributions with arbitrary shape parameter. The next remark shows a relation between the shape parameter ξ and the power of contortion $s \in \mathbb{N}$ in definition 4.2.

Remark 4.5. Let $\hat{F}$ be a baseline distribution such that $1 - \hat{F}(x) = \hat{c}x^{-1/\hat{\xi}}$, $x \ge x_q$, $\hat{c} > 0$ and $\hat{\xi} \in (0, 1)$. Let $s \in \mathbb{N}$ with $\hat{\xi} \le 1/s$ and $\mathcal{P}_\epsilon(\hat{F})$ the corresponding ambiguity set w.r.t. $d_{s,q}$. Then any alternative distribution F in $\mathcal{P}_\epsilon(\hat{F})$ such that $1 - F(x) = cx^{-1/\xi}$, $c \in \mathbb{R}_+$, $\hat{\xi} \le \xi$ satisfies $\xi \le 1/s$.

To see this, let $\hat{\gamma} = 1/\hat{\xi} \ge 1$ and $\gamma = 1/\xi \ge 1$ be the tail indices of the baseline distribution $\hat{F}$ and the alternative distribution F, respectively. We focus on the case $1 \le \gamma \le \hat{\gamma}$ where the alternative distribution has a heavier tail than the baseline one. The contorted Wasserstein distance is

$$\begin{aligned}\mathrm{WD}_{d_{s,q}}(F, \hat{F}) &= \int_0^{x_q} |F(x) - G(x)|\, dx + \int_{x_q}^{\infty} |F(x) - G(x)| s(x/x_q)^{s-1}\, dx \\ &= \int_0^{x_q} |F(x) - \hat{F}(x)|\, dx + \int_{x_q}^{\infty} \left| csx_q^{1-s}x^{-\gamma/s} - \hat{c}sx_q^{1-s}x^{-\hat{\gamma}/s} \right| dx < \infty,\end{aligned}$$

The second term in the above relation is bounded in only two cases.

Case 1. If $\hat{\xi} < 1/s$, then $\xi < 1/s$.

Case 2. If $\hat{\xi} = 1/s$, then $\xi = 1/s$ and $\hat{c} = c$.

Using the structure of the distributions as in assumption 4.1, the contorted Wasserstein distance between F and $\hat{F}$ is the area between the cumulative distribution functions F and $\hat{F}$ as in figure (3).

$$\begin{aligned}\mathrm{WD}_{d_{s,q}}(F, \hat{F}) &= \int_0^{x_q} \left|F(x) - \hat{F}(x)\right| dx + \int_{x_q}^{\infty} \left|F(x) - \hat{F}(x)\right| sx^{s-1}x_q^{1-s}\, dx \qquad (14) \\ &= \frac{1}{2}\sum_{i=1}^{n-1} \left(x^{(i+1)} - x^{(i)}\right) H\left(F^{(i)} - \hat{F}^{(i)}, F^{(i+1)} - \hat{F}^{(i+1)}\right) + \int_{x_q}^{\infty} |cx^{-1/s\xi} - \hat{c}x^{-1/s\hat{\xi}}| sx_q^{1-s}\, dx,\end{aligned}$$

where the function H computes the area of the trapezoid with corners $(\hat{F}^{(i)}, \hat{F}^{(i+1)}, F^{(i)}, F^{(i+1)})$, i.e.,

$$H(x, y) = \begin{cases} |x - y|, & \text{if } xy \ge 0 \\ (x^2 + y^2)/|x - y|, & \text{if } xy < 0. \end{cases}$$

The function H is convex in $F^{(i)}$, $i = 1, \dots, n$ (see Pflug et al. (2017), appendix). Contorted Wasserstein distance $\mathrm{WD}_{d_{s,q}}$ is linear in c and increasing in ξ. The computation of the alternative model F is shown in the appendix.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Problem (P_5) is solved using an iterative procedure (see algorithm 1). Note that according to proposition 3.2, the admissible contracts are the entrance excess-of-loss contracts characterized by parameters $0 \leq d_1 \leq d_2$. To emphasize the dependence on d_1 and d_2, we denote the contract by $I_d(x) := I(x) = \min(\max(x - d_1, 0), d_2 - d_1)$, where $d = (d_1, d_2)$.

Algorithm 1: Algorithm to solve problem (P_5).

Data: • Baseline model $\hat{F}$ satisfies assumption 4.1;
• Ambiguity radius $\epsilon > 0$;
• Number m of iterations;
• Set of considered models $\mathcal{C}$, $i = 1$.
Result: Optimal d_1^* and d_2^* such that $I_d(x) = \min(\max(x - d_1^*, 0), d_2^* - d_1^*)$ is insensitive w.r.t. $\mathcal{C}$.
initialization: $\mathcal{C} = \{\hat{F}\}$;
while $i \leq m$ **do**

Outer problem:
Input: $\mathcal{C}$;

$$\begin{aligned} \min_{d_1,d_2} \max_{F \in \mathcal{C}} \quad & \mathrm{AV@R}_\alpha(X^F - I_d(X^F) + \pi^{g,\theta}(I_d(X^{\hat{F}}))) \\ \text{s.t.} \quad & \pi^{g,\theta}(I_d(X^{\hat{F}})) \leq B \\ & 0 \leq d_1 \leq d_2 \end{aligned}$$

Output: parameters $d = (d_1^{(i)}, d_2^{(i)})$ of the contract and worst-case distribution $F^* \in \mathcal{C}$.
Inner problem:
Input: parameters $(d_1^{(i)}, d_2^{(i)})$ from outer problem;

$$\begin{aligned} \max_{F} \quad & \mathrm{AV@R}_\alpha(X^F - I_d(X^F) + \pi^{g,\theta}(I_d(X^{\hat{F}}))) \\ \text{s.t.} \quad & F \text{ satisfies assumption 4.1} \\ & \mathrm{WD}_{d_s,q}(F, \hat{F}) \leq \epsilon \end{aligned}$$

Output: alternative model F_i; update $\mathcal{C} = conv(\hat{F}, F_2, \ldots, F_{i-1}, F_i)$; $i = i + 1$.

end

The resulting worst-case distribution $F^* \in \mathcal{C} = conv(F_1, \ldots, F_m)$ for some alternative models $F_1, \ldots, F_m \in \mathcal{P}_\epsilon(\hat{F})$ and for some $\epsilon > 0$ will have a shape parameter ξ^* given by

$$\xi^* = \max_{i=\overline{1,m}} \{\xi_i | \, \lambda_i > 0\},$$

where $F^* = \sum_{i=1}^m \lambda_i F_i$, $\lambda_i \geq 0$, $\sum_{i=1}^m \lambda_i = 1$ and ξ_i is the shape parameter of F_i, $i = 1, \ldots, m$.

5 Numerical example

Tornadoes are extreme natural events that affect the U.S. mainland more than other parts of the world, with an annual average of 1200 events. The area on the east side of the Rocky Mountains, including parts of Oklahoma, Kansas, and northern Texas, is most prone to tornadoes, which is why it has received the name "tornado alley". Tornadoes of category F5 on the Fujita scale are considered extreme events, even though less than 1% of the total number fall into this category, but may cause significant damage. Consequently, there is an increasing need for more efficient tools in risk assessment and insurance mechanisms in the face of such extreme events.

Data is taken from the Storm Events Database (*https://www.ncdc.noaa.gov/stormevents*), which contains records created by the official United States National Oceanic and Atmospheric Administration (NOAA). Of all the meteorological events registered, we focus on tornadoes, since this type of event has the longest period of record, i.e., 1951-2015. Each tornado is coded as an episode which may contain one or more events, uniquely identified by a key. For each such event, there are around 50 variables which include, among others, the state affected by the particular tornado, the date of the beginning and end of the phenomenon, its length and width while on the ground, the number of dead or injured people (directly or indirectly) and its F-scale. The direct economic losses caused by tornadoes include property and crop damage, determined in the weeks and sometimes months after the event. The indirect damage (long-term macroeconomic effects and loss of human life) are excluded.

ACCEPTED MANUSCRIPT

The claims included in this database are gathered from insurance companies, mass-media, or other non-official sources, and the data we use may thus already contain some uncertainty. Moreover, especially for extreme tornadoes, the lack of data and the difficulty in forecasting together increase the model ambiguity. These sources of ambiguity enable us to apply the framework developed in the previous sections.

The losses provided by NOAA are first adjusted for inflation in 2015 dollars and then rescaled in billions of dollars. The first step in our procedure is the estimation of a baseline distribution $\hat{F}$ of losses. Based on data and using statistical tools from EVT, a variety of distributions belonging to the class of GEV and GPD are tested and the goodness of fit is verified using graphical tests such as P-P plots and Q-Q plots (see Coles et al. (2001)). From this analysis, the baseline distribution is considered as GPD with a shape parameter $\xi = 0.45$. However, the choice of an appropriate threshold is a crucial first step in fitting GPD: on one hand, the threshold must be sufficiently high to ensure the asymptotic behaviour of GPD and on other, be low enough to allow parameter estimation (for a review see Scarrott and MacDonald (2012)). This situation of epistemic uncertainty entitles the use of an ambiguity set in the design of optimal insurance contracts.

The minimax optimization problem is solved according to algorithm 1. As input, the level at which the tails of distributions are assumed to start is $q = 0.997$ and the power of contortion on $\mathbb{R}$ is $s = 2$. To compute the premium, we employed the distortion power $g(x) = x^{0.5}$, and the level for AV@R$_\alpha$ is $\alpha = 0.8$. The budget for the premium is $B = 1.2$ (in billions of dollars) and parameter $\theta = 0.2$. If the ambiguity radius is considered to be $\epsilon = 0.5$, then the optimization problem to solve is the following:

$$\begin{aligned} \min_{d_1,d_2} \max_{F} \quad & \mathrm{AV@R}_{0.8}\big(X^F - I_d(X^F)\big) + \pi^{g,\theta}(I_d(X^{\hat{F}})) \\ \text{subject to} \quad & \pi^{g,\theta}(I_d(X^{\hat{F}})) \le 1.2 \\ & \mathrm{WD}_{d_{2,0.997}}(F, \hat{F}) \le 0.5 \\ & 0 \le d_1 \le d_2. \end{aligned}$$

When the optimal value is reached, as already obtained in proposition 3.2, the premium w.r.t. the baseline distribution equals the available budget. The optimal values of the parameters are $d_1 = 0.5092$ and $d_2 = 3.0879$. For this input, the premium calculated w.r.t. the worst-case model is 1.242.

We also solved the problem for different ambiguity radii and studied the dependence of the objective function as well as the deductible and cap levels of the insurance contract on the tolerance level change. As we can observe from figure (4), both parameters of this contract are increasing with the increase in the ambiguity radius. In the risk-averse setting, the insured person is more likely to cover the small losses using a risk reduction procedure, in exchange for protection against high losses offered by the insurance company.

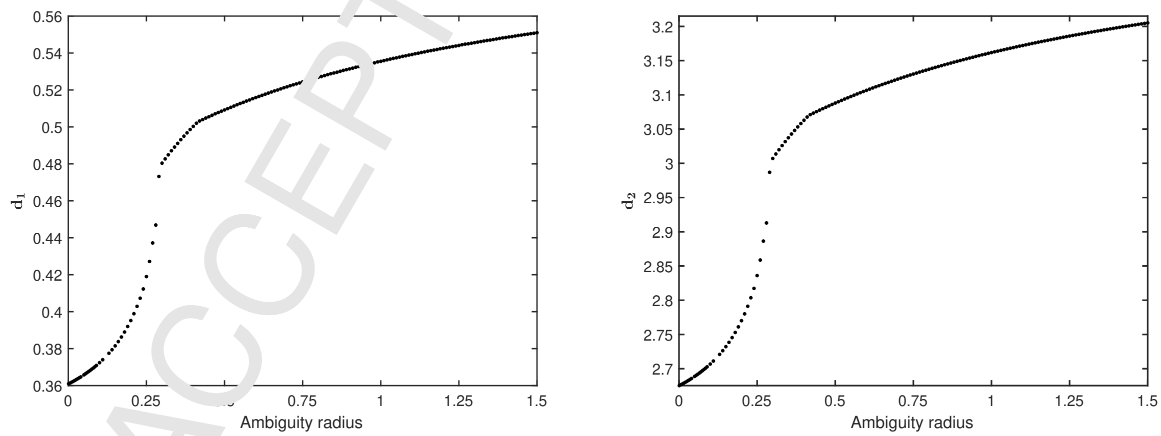


Figure 4: Dependence of d_1 (left) and d_2 (right) on ϵ.

We define the *ambiguity premium* as the difference between the insurance premium under ambiguity and the insurance premium computed w.r.t. the baseline distribution. More precisely,

$$\pi_{\text{ambiguity}} = \max_{F \in \mathcal{C}} \pi^{g,\theta}(I_d(X^F)) - \pi^{g,\theta}(I_d(X^{\hat{F}})).$$

ACCEPTED MANUSCRIPT

As expected, an increase in the ambiguity set results in an increase in this difference (see table (5)).

Table 1: Optimal parameters of XL-contract and premium w.r.t. worst-case model.

ϵ	d_1	d_2	$\pi^{g,\theta}(X^{F^*})$	$\pi_{\text{ambiguity}}$
0	0.3609	2.6750	1.2000	0
0.006	0.3614	2.6765	1.2005	$5 \cdot 10^{-4}$
0.04	0.3647	2.6857	1.2031	$3.1 \cdot 10^{-3}$
0.08	0.3695	2.6989	1.2060	$6 \cdot 10^{-3}$
0.2	0.3952	2.7701	1.2145	$1.43 \cdot 10^{-2}$
0.7	0.5217	3.1230	1.2585	$5.85 \cdot 10^{-2}$
1	0.5355	3.1618	1.2819	$8.19 \cdot 10^{-2}$
1.3	0.5456	3.1900	1.3043	0.10[illegible]3
1.5	0.5510	3.2051	1.3191	0.1191

The size of the ambiguity radius depends on the amount of information available and hence, on the risk-averse attitude of the participants in the insurance market. Typically, a larger sample size allows the size of the ambiguity set to be decreased.

At each change of ϵ, the minimax problem yields a worst-case distribution $F^* \in \mathcal{C}$ which is a convex combination of alternative models $F_1, \ldots, F_m$ (see figure (6) a). The risk and premium corresponding to the worst-case model are illustrated in figure (5). As expected, both quantities increase with the enlargement of the ambiguity radius, emphasizing the effect of model ambiguity on risk assessment and insurance premium.

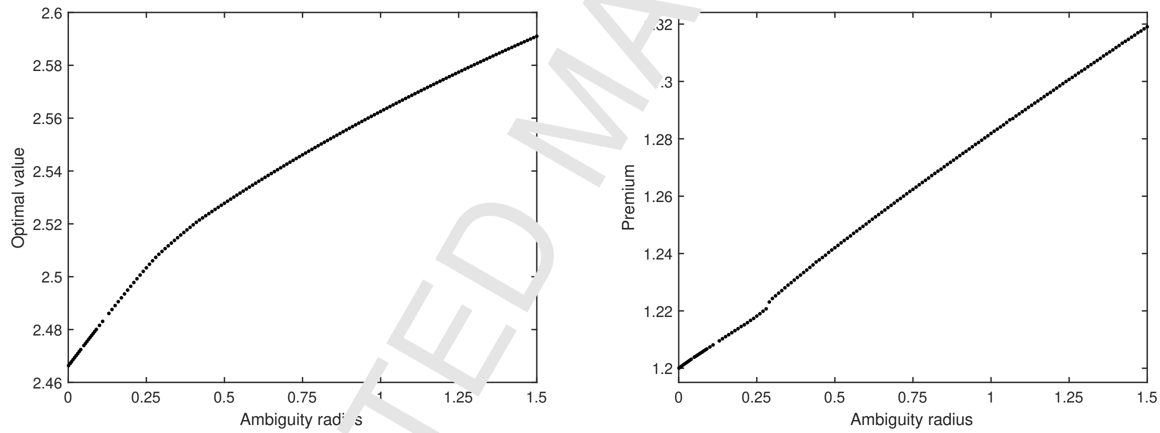


Figure 5: Dependence of risk and π on ϵ.

The worst-case distributions associated with the minimal risk in figure (5) a) are shown in figure (6) a). The shape parameter of the worst-case distribution increases with the size of the ambiguity radius, generating heavier-tailed distributions. From remark 4.5, the shape parameter of each F^* in algorithm 1 is between 0.45 and 0.5 (see figure (6), b)); however, the upper limit is obtained for large values of ambiguity radius.

To minimize the risk of the retained losses, the insurance contract would cover more of the extreme claims, which are more expensive for the insured person. Therefore, for small ϵ, the worst-case distribution is close to the baseline distribution, and there is a steep increase in the values of d_1 and d_2, from 0.36 to 0.48 and from 2.6 to 3, respectively, for $\epsilon \in [0, 0.3]$. For larger values of ϵ, the worst-case distribution stochastically dominates the baseline distribution; therefore, covering large losses requires a significant increase in the premium. In this case, the parameters d_1 and d_2 are increasing at a slower rate, i.e., $d_1 \in (0.48, 0.58)$ and $d_2 \in (3, 3.2)$ over a range of ϵ from 0.3 to 1.5. The value of ϵ at which this change in behavior happens depends on the choice of $x^{(i)}$, $i = \overline{1, n}$ used to construct alternative models in the Wasserstein ball.

From the decision process point of view, it is advisable to maintain part of the budget for an increase in the premium to protect against possible model misspecification.

ACCEPTED MANUSCRIPT

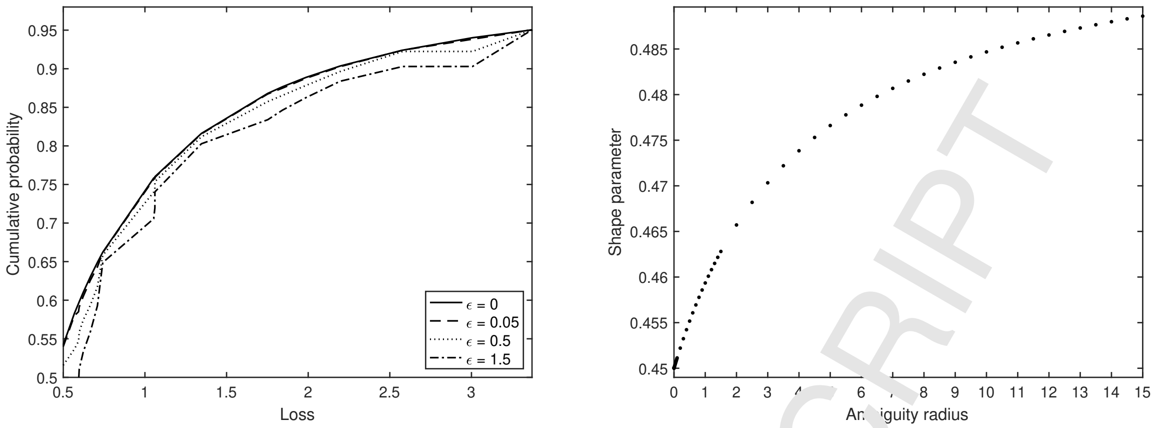


Figure 6: a) Worst-case distributions for different ϵ (closer look). b) Dependence of ξ on ϵ.

6 Conclusion

The classical approach for designing an optimal insurance contract relies on the assumption that the loss distribution is completely known. However, estimation errors or lack of information can lead to uncertainty about a single suitable model. The model ambiguity increases even more when dealing with extreme natural events due to the limited number of observations and the global dynamics typical of rare events. Considering these sources of ambiguity, our aim in this paper is to determine an insurance contract which is robust under possible model misspecification. Through a stochastic optimization approach, we study the optimal balance between the contract parameters that minimize some risk functional of the retained loss. To include model ambiguity, a set of feasible models is incorporated into the decision process, resulting in a minimax formulation. This set is constructed based on a modified version of the Wasserstein distance, which is more appropriate for heavy-tailed distributions. The resulting solution proves robust in the following sense: this insurance contract might be slightly sub-optimal w.r.t the baseline model, but it is stable under models within the ambiguity set of the base model. Sensitivity analysis and numerical implementations are addressed, and the performance is assessed using an insurance claims dataset.

Acknowledgments

The authors are grateful for the comments from two anonymous referees, which have led to the improvement of the present paper. The work was partially supported by the Vienna Graduate School on Computational Optimization (VGSCO)(FWF-Project No. W1260-N35) (C. Birghila) and International Institute for Applied Systems Analysis (IIASA)(G. Pflug).

References

Alary, D., Gollier, C., and Treich, N. (2013). The effect of ambiguity aversion on insurance and self-protection. *The Economic Journal*, 123(573):1188–1202.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American economic review*, 53(5):941–973.

Asimit, A. 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.

Assa, H. (2015). On optimal reinsurance policy with distortion risk measures and premiums. *Insurance: Mathematics and Economics*, 61:70–75.

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

ACCEPTED MANUSCRIPT

Balkema, A. A. and Haan, L. d. (1974). Residual life time at great age. *The Annals of probability*, pages 792–804.

Bernard, C. and Tian, W. (2009). Optimal reinsurance arrangements under tail risk measures. *Journal of risk and insurance*, 76(3):709–725.

Borch, K. (1960). An attempt to determine the optimum amount of stop loss reinsurance. *Transactions of the 16th International Congress of Actuaries*, I:597–610.

Cheung, K., Liu, F., and Yam, S. (2012). Average Value-at-Risk minimizing reinsurance under Wang's premium principle with constraints. *ASTIN Bulletin: The Journal of the IAA*, 42(2):575–600.

Chi, Y. and Tan, K. S. (2011). Optimal reinsurance under Var and CVaR risk measures: a simplified approach. *ASTIN Bulletin: The Journal of the IAA*, 41(2):487–509.

Chi, Y. and Tan, K. S. (2013). Optimal reinsurance with general premium principles. *Insurance: Mathematics and Economics*, 52(2):180–189.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.

Cummins, J. D. and Mahul, O. (2004). The demand for insurance with an upper limit on coverage. *Journal of Risk and Insurance*, 71(2):253–264.

Denneberg, D. (1990). Distorted probabilities and insurance premiums. *Methods of Operations Research*, 63(3).

Dhaene, J., Kukush, A., Linders, D., and Tang, Q. (2012). Remarks on quantiles and distortion risk measures. *European Actuarial Journal*, 2(2):319–328.

Doherty, N. A., Laux, C., and Muermann, A. (2015). Insuring Nonverifiable Losses. *Review of Finance*, 19(1):283–316.

Dudley, R. (1969). The speed of mean Glivenko-Cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, pages 643–669.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media.

Esfahani, P. M. and Kuhn, D. (2017). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, pages 1–52.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24(2), pages 180–190. Cambridge University Press.

Gnedenko, B. (1943). Sur la distribution limité du terme maximum d'une série aléatoire. *Annals of mathematics*, pages 423–453.

Gollier, C. (2014). Optimal insurance design of ambiguous risks. *Economic Theory*, 57(3):555–576.

Goovaerts, M. J., Kaas, R., and Laeven, R. J. (2011). Worst case risk measurement: Back to the future? *Insurance: Mathematics and Economics*, 49(3):380–392.

Kaluszka, M. (2001). Optimal reinsurance under mean-variance premium principles. *Insurance: Mathematics and Economics*, 28(1):61–67.

Kiesel, R., Rühlicke, R., Stahl, G., and Zheng, J. (2016). The Wasserstein metric and robustness in risk management. *Risks*, 4(3):32.

Klibanoff, P., Marinacci, M., and Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892.

ACCEPTED MANUSCRIPT

Klibanoff, P., Marinacci, M., and Mukerji, S. (2009). Recursive smooth ambiguity preferences. *Journal of Economic Theory*, 144(3):930–976.

Lo, A. (2017a). A Neyman-Pearson perspective on optimal reinsurance with constraints. *ASTIN Bulletin: The Journal of the IAA*, 47(2):467–499.

Lo, A. (2017b). A unifying approach to risk-measure-based optimal reinsurance problems with practical constraints. *Scandinavian Actuarial Journal*, 2017(7):584–605.

Pflug, G. and Wozabal, D. (2007). Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442.

Pflug, G. C. and Pichler, A. (2016). *Multistage stochastic optimization*. Springer.

Pflug, G. C., Timonina-Farkas, A., and Hochrainer-Stigler, S. (2017). Incorporating model uncertainty into optimal insurance contract design. *Insurance: Mathematics and Economics*, 73:68–74.

Pickands, J. et al. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1):119–131.

Prokhorov, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications*, 1(2):157–214.

Raviv, A. (1979). The design of an optimal insurance policy. *The American Economic Review*, 69(1):84–96.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

Sion, M. et al. (1958). On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176.

Tan, K. S., Weng, C., and Zhang, Y. (2011). Optimality of general reinsurance contracts under CTE risk measure. *Insurance: Mathematics and Economics*, 49(2):175–187.

Vallender, S. (1974). Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications*, 18(4):784–786.

Villani, C. (2008). *Optimal transport: old and new*, volume 338. Springer Science & Business Media.

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

Young, V. R. (1999). Optimal insurance under Wang's premium principle. *Insurance: Mathematics and Economics*, 25(2):109–122.

Zhuang, S. C., Weng, C., Tan, K. S., and Assa, H. (2016). Marginal indemnification function formulation for optimal reinsurance. *Insurance: Mathematics and Economics*, 67:65–76.

Zymler, S., Kuhn, D., and Rustem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198.

Appendix

Proof of Remark 4.1. 2. When $x_q \geq K$, then $d_{s,x_q} = d_1$. So let $1 << x_q < K$.

Case 1. If $x, y < x_q$, then $\varphi_{s,x_q}(x) = x$ and $|x-y| = |\varphi_{s,x_q}(x) - \varphi_{s,x_q}(y)|$.

Case 2. If $x_q \leq x, y \leq K$, then $\varphi_{s,x_q}(x) = x_q^{1-s}x^s$ and $|x_q^{1-s}x^s - x_q^{1-s}y^s| = x_q^{1-s}|x-y|(x^{s-1}+\ldots+y^{s-1})$ which is bounded by

$$x_q^{1-s}|x-y| \leq x_q^{1-s}|x^s - y^s| \leq x_q^{1-s}sK^{s-1}|x-y|.$$

Case 3. If $x \leq x_q \leq y \leq K$, then $\varphi_{s,x_q}(x) = x$ and $\varphi_{s,x_q}(y) = x_q^{1-s}y^s$. Therefore $|x - x_q^{1-s}y^s| \geq x_q^{1-s}y^s - x \geq |x-y|$.
Since $x < x_q$, then $x_q^{1-s}x^s \leq x$ and hence $|x_q^{1-s}y^s - x| \leq |x_q^{1-s}y^s - x_q^{1-s}x^s| \leq x_q^{1-s}sK^{s-1}|x-y|$.

ACCEPTED MANUSCRIPT

Therefore $\exists\, C_1, C_2 > 0$ constants such that $C_1|x-y| \leq d_{s,x_q}(x,y) \leq C_2|x-y|$.

$\square$

Definition 6.1 (Renewal model)**.** The renewal model is given by the following conditions:

(a) The *claim size process*: the claim sizes $(X_i)_{i\in\mathbb{N}} \overset{iid}{\sim} F$ are positive random variables, $\mathbb{E}(X_i^F) = \mu < \infty$;

(b) The *claim times*: the claims occur at the random instants of time $0 < T_1 < T_2 < \ldots$ a.s.;

(c) The *claim arrival process*: the number of claims in the interval $[0,t]$ is denoted by

$$N(t) := \sup\{n \geq 1 : T_n \leq t\};$$

(d) The inter-arrival times: $(Y_i)_{i\in\mathbb{N}}$ are iid random variables with $\mathbb{E}(Y_i) = 1/\lambda$, where Y_i is defined as

$$Y_1 = T_1, Y_i = T_i - T_{i-1}, i = 2, 3, \ldots;$$

(e) $(X_i)_{i\in\mathbb{N}}$ and $(Y_i)_{i\in\mathbb{N}}$ are independent of each other.

For simplicity, we denoted the renewal process by the couple $\big(X_i, Y_i\big)$.

Definition 6.2 (Ruin probability)**.** The *ruin probability in infinite time* is defined as

$$\psi(u,F) = P\bigg(u + ct - \sum_{i=1}^{N(t)} X_i < 0, \text{ for some } t \leq \infty\bigg),$$

where $u \geq 0$ the *initial capital* and $c > 0$ the *premium income rate*.

Remark 6.1. Let F be the distribution of claim sizes $(X_i)_{i\in\mathbb{N}}$ as in definition 6.1. If the survival distribution $\overline{F}(x) = 1 - F(x)$ is regular varying of some index γ, then for large enough capital u, the following holds

$$\lim_{u\to\infty} \frac{1}{\psi(u,F)} \cdot \frac{\lambda}{c - \lambda\mu} \int_u^\infty \overline{F}(x)\, dx = 1. \tag{15}$$

For a proof, see Embrechts et al. (2013), Chapter 1.

Proof of proposition 4.1 3. Denote $\mu_1 = \mathbb{E}(X_i^F)$ and $\mu_2 = \mathbb{E}(\tilde{X}_i^{\,G})$. From remark 6.1, there exists some $\delta_1, \delta_2 \geq 0$ such that $\psi(u,F) \leq (1+\delta_1)\dfrac{\lambda}{c-\lambda\mu_1}\displaystyle\int_u^\infty \overline{F}(x)\,dx$ and $\psi(u,G) \leq (1+\delta_2)\dfrac{\lambda}{c-\lambda\mu_2}\displaystyle\int_u^\infty \overline{G}(x)\,dx$. If $\delta := \max\{\delta_1, \delta_2\}$,

$$|\psi(u,F) - \psi(u,G)| \leq (1+\delta)\lambda \int_u^\infty \Big|\frac{1}{c-\lambda\mu_1}\overline{F}(y) - \frac{1}{c-\lambda\mu_2}\overline{G}(x)\Big|\, dy \tag{16}$$

Denote $a := \dfrac{1}{c-\lambda\mu_1}$ and $b := \dfrac{1}{c-\lambda\mu_2}$. For all $x \in (u,\infty)$, the integrand becomes

$$|a\overline{F}(x) - b\overline{G}(x)| = |a\overline{F}(x) - b\overline{F}(x) + b\overline{F}(x) - b\overline{G}(x)| \leq |a-b|\overline{F}(x) + b|\overline{F}(x) - \overline{G}(x)|$$
$$|a\overline{F}(x) - b\overline{G}(x)| = |a\overline{F}(x) - a\overline{G}(x) + a\overline{G}(x) - b\overline{G}(x)| \leq a|\overline{F}(x) - \overline{G}(x)| + |a-b|\overline{G}(x)$$

Summing up the two inequalities, we obtain

$$|a\overline{F}(x) - b\overline{G}(x)| \leq \frac{1}{2}\Big(|a-b|\big(\overline{F}(x) + \overline{G}(x)\big) + (a+b)|\overline{F}(x) - \overline{G}(x)|\Big).$$

Then (16) becomes

$$\begin{aligned}
|\psi(u,F) - \psi(u,G)| &\leq \frac{(1+\delta)\lambda}{2}\Big[|a-b|\Big(\int_u^\infty \overline{F}(x)dx + \int_u^\infty \overline{G}(x)dx + \Big) + (a+b)\int_u^\infty |\overline{F}(x) - \overline{G}(x)|\,dx\Big] \\
&\leq \frac{(1+\delta)\lambda}{2}\Big[|a-b|(\mu_1+\mu_2) + (a+b)\int_0^\infty |\overline{F}(x) - \overline{G}(x)|\,dx\Big] \\
&\leq \frac{(1+\delta)\lambda}{2}\Big[|a-b|(\mu_1+\mu_2) + (a+b)\int_0^\infty |\overline{F}(x) - \overline{G}(x)|\,d\varphi_{s,q}(x)\Big] \\
&= \frac{(1+\delta)\lambda}{2}\Big[\frac{|\mu_1-\mu_2|}{(c-\lambda\mu_1)(c-\lambda\mu_2)}\lambda(\mu_1+\mu_2) + (a+b)\mathrm{WD}_{d_{s,x_q}}(F,G)\Big] \\
&\leq \frac{(1+\delta)\lambda}{2}\Big[\frac{\lambda(\mu_1+\mu_2)}{(c-\lambda\mu_1)(c-\lambda\mu_2)}\mathrm{WD}_{d_{s,x_q}}(F,G) + (a+b)\mathrm{WD}_{d_{s,x_q}}(F,G)\Big] \\
&= C \cdot \mathrm{WD}_{d_{s,x_q}}(F,G).
\end{aligned}$$

ACCEPTED MANUSCRIPT

For $\delta \to 0$, $C = \frac{\lambda c}{(c-\lambda\mu_1)(c-\lambda\mu_2)} > 0$. □

Computation of alternative model F. We consider that distributions in the contorted Wasserstein ball follow assumption 4.1. From the computational point of view, each distribution F in $\mathcal{P}_\epsilon(\hat{F})$, for some $\epsilon > 0$ is an n-tuple given by $(F^{(1)}, F^{(2)}, \dots, F^{(n-1)}, c, \xi)$ such that $F^{(0)} = 0$ and $F^{(n)} = \hat{F}^{(n)}$. Using these approximations of the distributions, we have an analytical expression to compute the contorted Wasserstein distance between F and $\hat{F}$ given by (14). The inner optimization problem then becomes:

$$\begin{aligned}
&\max_{\boldsymbol{v},c,\xi} \mathrm{AV@R}_\alpha(X^F - I(X^F) + \pi^{g,\theta}(I(X^{\hat{F}})))\\
&\text{s.t. } A\boldsymbol{v} \le \mathbf{0}\\
&\quad \mathbf{0} \le \boldsymbol{v} \le \mathbf{1}\\
&\quad \mathrm{WD}_{d_{s,q}}(F, \hat{F}) \le \epsilon,
\end{aligned}$$

where $A = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(n-2)\times(n-1)}$ and $\boldsymbol{v} = (F^{(1)}, F^{(2)}, \dots, F^{(n-1)})^\top$.

Saddle point and minimax theorem.

Let $\mathbb{X}$ and $\mathbb{Y}$ be two topological spaces and let f be a real valued function on $\mathbb{X} \times \mathbb{Y}$. We consider the minimax problem

$$\min_{x\in\mathbb{X}} \min_{y\in\mathbb{Y}} f(x,y) \tag{P}$$

Notice first that

$$\min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y) \ge \max_{y\in\mathbb{Y}} \min_{x\in\mathbb{X}} f(x,y)$$

holds always. The reverse inequality needs the existence of a saddle point.

Definition 6.3. A pair (x^*, y^*) is called *saddle point* of f if

$$f(x^*, y) \le f(x^*, y^*) \le f(x, y^*)$$

for all $x \in \mathbb{X}$ and $y \in \mathbb{Y}$.

Observe now that if (x^*, y^*) is a saddle point, it follows that

$$\min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y) \le \max_{y\in\mathbb{Y}} f(x^*, y) = f(x^*, y^*) = \min_{x\in\mathbb{X}} f(x, y^*) \le \max_{y\in\mathbb{Y}} \min_{x\in\mathbb{Y}} f(x,y).$$

The relation between saddle points and the solutions of the problem (P) requires additional assumptions on the structure of spaces $\mathbb{X}$ and $\mathbb{Y}$, as well as on the function f.

Theorem 6.1 (Sion et al. (1958))**.** Let $\mathbb{X}$ and $\mathbb{Y}$ be two convex subsets of a linear topological space. Suppose that f is a quasi convex-concave function on $\mathbb{X} \times \mathbb{Y}$, such that $f(x,\cdot)$ is upper semi-continuous on $\mathbb{Y}$, for all $x \in \mathbb{X}$, and $f(\cdot, y)$ is lower semi-continuous on $\mathbb{X}$, for all $y \in \mathbb{Y}$. If $\mathbb{X}$ is compact, then

$$\min_{x\in\mathbb{X}} \sup_{y\in\mathbb{Y}} f(x,y) = \sup_{y\in\mathbb{Y}} \min_{x\in\mathbb{X}} f(x,y).$$

If $\mathbb{Y}$ is compact, then

$$\inf_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y) = \max_{y\in\mathbb{Y}} \inf_{x\in\mathbb{X}} f(x,y).$$

If both are compact, then

$$\min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y) = \max_{y\in\mathbb{Y}} \min_{x\in\mathbb{X}} f(x,y).$$

We mention another important result required in the proof of proposition 3.1.

Theorem 6.2 (Prokhorov (1956))**.** Let $(\mathcal{X}, d)$ be a complete separable metric space, $\mathcal{P}(\mathcal{X})$ the set of all Borel probability measures on $\mathcal{X}$ and $\mathcal{Y}$ be a subset of $\mathcal{P}(\mathcal{X})$. Then $\mathcal{Y}$ is tight if and only if the closure $\overline{\mathcal{Y}}$ of $\mathcal{Y}$ is compact in $\mathcal{X}$.