International Journal of Approximate Reasoning 161 (2023) 108986

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Addressing ambiguity in randomized reinsurance stop-loss treaties using belief functions

Davide Petturiti [a,*], Gabriele Stabile [b], Barbara Vantaggi [b]

[a] *Dip. Economia, Università degli Studi di Perugia, 06123 Perugia, Italy*
[b] *Dip. MEMOTEF, "La Sapienza" Università di Roma, 00161 Roma, Italy*

ARTICLE INFO

Article history:
Received 31 March 2023
Received in revised form 14 July 2023
Accepted 14 July 2023
Available online 20 July 2023

Keywords:
Belief function
Dempster-Shafer approximation
Reinsurance
Stop-loss treaty

ABSTRACT

The aim of the paper is to model ambiguity in a randomized reinsurance stop-loss treaty. For this, we consider the lower envelope of the set of bivariate joint probability distributions having a precise discrete marginal and an ambiguous Bernoulli marginal. Under an independence assumption, since the lower envelope fails 2-monotonicity, inner/outer Dempster-Shafer approximations are considered, so as to select the optimal retention level by maximizing the lower expected insurer's annual profit under reinsurance. We show that the inner approximation is not suitable in the reinsurance problem, while the outer approximation preserves the given marginal information, weakens the independence assumption, and does not introduce spurious information in the retention level selection problem. Finally, we provide a characterization of the optimal retention level.

1. Introduction

Uncertainty is usually modeled through a probability measure, however a demand for more flexible models arises in different fields with the aim to provide tools able to manage partially specified information (imprecision) through a class of compatible probability measures. This work is essentially motivated by an application related to reinsurance and aims at incorporating ambiguity in a simple randomized reinsurance model, introduced in [2].

A *reinsurance treaty* (see, e.g., [1]) is a contract between a *reinsurer* and a *(first-line) insurer* (also called *cedent*) for transferring some parts of the insurance risk, upon the payment of a *reinsurance premium*. Most of non-life reinsurance contracts are actually written for a one year horizon, thus we will focus on a one-year period. Given a non-negative random variable X, expressing the aggregate loss that the insurer faces over the one-year period, a *stop-loss treaty* is a contract in which the retained loss of the insurer is the random variable $r(X,d) = \min(X,d)$, where $d \geq 0$ is a fixed *retention level*.

In [2], the classical stop-loss treaty is randomized according to an independent Bernoulli random variable Y with parameter p, leading to a retained loss $r(X,Y,d)$ which is $\min(X,d)$ if $Y = 1$ and X otherwise. In such a randomization, the parameter p of the Bernoulli distribution can be interpreted as the probability of a rare event or as a default probability. For a fixed p, the goal is to choose a retention level d^*, that maximizes the expected annual profit $Z(X,Y,d)$, by taking into account the total premium $\pi(X)$ received from the first-line insurer, the reinsurance premium $\pi_R(d)$, the cost-of-capital rate r_{coc}, and a solvency risk measure $\rho(r(X,Y,d))$.

* Corresponding author.
E-mail addresses: davide.petturiti@unipg.it (D. Petturiti), gabriele.stabile@uniroma1.it (G. Stabile), barbara.vantaggi@uniroma1.it (B. Vantaggi).

https://doi.org/10.1016/j.ijar.2023.108986

Since the parameter p can suffer from misspecification, the aim of present paper is to investigate the effect of ambiguity on the choice of the optimal retention level d. A random vector (X, Y) is considered with X having a discrete distribution P_X, whereas the distribution of Y belongs to a class $\mathcal{P}_Y$ of Bernoulli distributions, where p ranges in a closed interval. Under the hypothesis of independence of the two variables, meaning that under P_X and any element of $\mathcal{P}_Y$ the two variables are independent, we prove that the class of joint probability distributions generates a class of probability measures (credal set) $\mathcal{P}$, that is closed and convex, but whose lower envelope $\underline{P}$ is generally not 2-monotone. Moreover, the core of $\underline{P}$ (i.e., the set of all probability measures dominating $\underline{P}$) strictly contains $\mathcal{P}$.

Following the approach of [21], we first look for an inner Dempster-Shafer approximation Bel^i of $\underline{P}$, that is a belief function dominating $\underline{P}$ and minimizing the Euclidean distance. Depending on the chosen parameters, such Bel^i can be either additive or non-additive, but its use in the reinsurance application is impaired since between the core of Bel^i and the class $\mathcal{P}$ no containment relationship may hold.

For this, we move towards an outer Dempster-Shafer approximation [22,25,26] that further preserves the marginal probability distribution of X, namely Bel^{oo}. Such a belief function is asked to dominate $\underline{P}$, to preserve the marginal probability distribution P_X of X and to minimize the Euclidean distance. We provide a closed-form expression for Bel^{oo} and show that Bel^{oo} also preserves the imprecise marginal of Y, that is $\underline{P}_Y = \min \mathcal{P}_Y$. Since $\mathcal{P}$ is strictly contained in the core of Bel^{oo}, its use allows us to weaken the independence assumption between X and Y which is, in turn, not easy to justify when interpreting p as the probability of a rare event or a default probability.

Both Bel^i and Bel^{oo} allow the computation of lower expectations (with respect to their core) as Choquet expectations and are used to model the lower expected insurer's annual profit under reinsurance. The use of the Choquet integral, with respect to either Bel^i or Bel^{oo}, subsumes a *maximin* criterion of choice, nevertheless, besides the Choquet integral, other expectation operators could be considered inside Dempster-Shafer theory [35]: this will be addressed in future research. Incidentally, Bel^i and Bel^{oo} are proved to be optimal approximations of $\underline{P}$ in the wider framework of 2-monotone lower probabilities.

Following the approach in [2], for the solvency risk measure $\rho(r(X, Y, d))$ we adopt an ambiguous version of the value-at-risk (see, e.g., [19] for the classical definition) that relies on the lower cumulative distribution function induced by Bel^i or Bel^{oo}. Such choice is motivated by the common use of the value-at-risk in solvency contexts, joined with a cautious approach towards ambiguity due to the lower envelopes Bel^i and Bel^{oo}.

We show that the lower expected profit computed with respect to $\mathcal{P}$ coincides with the Choquet expected profit computed with respect to the outer approximation Bel^{oo}, thus their maximization to determine the optimal retention level d^* gives rise to the same optimization problem. In other terms, by working with Bel^{oo} we weaken the independence assumption between X and Y, and further we do not introduce spurious information in the reinsurance optimization problem. This leads us to focus on Bel^{oo} only, for which we provide a characterization of the optimal d^*.

This paper extends some preliminary results presented in [31]. The paper is structured as follows. Section 2 collects the necessary preliminaries. Section 3 introduces the inner and outer Dempster-Shafer approximations and provides a characterization of the outer approximation, together with a study of its properties. Section 4 introduces ambiguity in a randomized stop-loss treaty and shows that the outer Dempster-Shafer approximation leads to the same retention level selection problem of the original lower joint probability. Section 5 provides a characterization of the optimal retention level under the outer Dempster-Shafer approximation. Finally, Section 6 collects our conclusions and future perspectives.

2. Preliminaries

Let $\Omega = \{\omega_1, \dots, \omega_n\}$ be a finite non-empty set and denote by 2^Ω its power set. A function $\underline{P} : 2^\Omega \to [0,1]$ such that $\underline{P}(\emptyset) = 0$ and $\underline{P}(\Omega) = 1$ is called a:

- *(coherent) lower probability* if there exists a closed set $\mathcal{P}$ of probability measures on 2^Ω such that, for every $A \in 2^\Omega$,

$$\underline{P}(A) = \min_{P \in \mathcal{P}} P(A);$$

- *k-monotone lower probability* with $k \geq 2$ if for every $A_1, \dots, A_k \in 2^\Omega$,

$$\underline{P}\left(\bigcup_{i=1}^k A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\dots,k\}} (-1)^{|I|-1} \underline{P}\left(\bigcap_{i \in I} A_i\right).$$

A lower probability which is k-monotone for every $k \geq 2$ is called a *belief function* and is denoted as Bel [13,34].

Every lower probability $\underline{P}$ induces the closed (in the product topology) convex set of probability measures on 2^Ω, called *core*, defined as

$$\mathbf{core}(\underline{P}) = \{P \;:\; P \text{ is a probability measure on } 2^\Omega,\, P \geq \underline{P}\}.$$

In general, there may be infinitely many closed and convex sets of probability measures inducing a lower probability $\underline{P}$: such sets are also referred to as *credal sets* [20]. The set $\mathbf{core}(\underline{P})$ is the largest credal set associated with $\underline{P}$.

A lower probability $\underline{P}$ is completely determined (see [18]) by its *Möbius inverse* $m : 2^\Omega \to \mathbb{R}$ through the relations, for all $A \in 2^\Omega$,

$$m(A) = \sum_{B\subseteq A}(-1)^{|A\setminus B|}\underline{P}(B) \quad \text{and} \quad \underline{P}(A) = \sum_{B\subseteq A} m(B). \tag{1}$$

The function m is such that $m(\emptyset) = 0$, $m(\{\omega_i\}) \geq 0$, for all $\omega_i \in \Omega$, and $\sum_{A\in 2^\Omega} m(A) = 1$. We have that $\underline{P}$ is 2-monotone (see [7]) if and only if m further satisfies

$$\sum_{\{\omega_i,\omega_j\}\subseteq B\subseteq A} m(B) \geq 0, \quad \text{for all } A \in 2^\Omega \text{ and all } \omega_i, \omega_j \in A, \omega_i \neq \omega_j, \tag{2}$$

while $\underline{P}$ is a belief function if and only if m ranges in $[0, 1]$ (see, e.g., [18,34]).

Denoting by $\mathbb{R}^\Omega$ the set of all random variables on Ω, the issue of introducing a notion of expectation with respect to a closed set of probability measures $\mathcal{P}$ can be faced in two different manners: either referring to the Choquet integral with respect to the lower probability $\underline{P}$ or to the lower expectation functional with respect to $\mathcal{P}$. Given $\underline{P}$ and $X \in \mathbb{R}^\Omega$, the *Choquet expectation* of X with respect to $\underline{P}$ (see, e.g., [14,18]) is defined through the Choquet integral

$$\mathbb{C}_{\underline{P}}[X] = \sum_{i=1}^{n}(X(\omega_{\sigma(i)}) - X(\omega_{\sigma(i+1)}))\underline{P}(E_i^\sigma),$$

where σ is a permutation of Ω such that $X(\omega_{\sigma(1)}) \geq \ldots \geq X(\omega_{\sigma(n)})$, $E_i^\sigma = \{\omega_{\sigma(1)}, \ldots, \omega_{\sigma(i)}\}$ for $i = 1, \ldots, n$, and $X(\omega_{\sigma(n+1)}) = 0$.

In particular, if $\underline{P}$ reduces to a probability measure P, then $\mathbb{C}_P[X] = \mathbb{E}_P[X]$, where $\mathbb{E}_P$ denotes the usual expectation operator with respect to P. On the other hand, given $\mathcal{P}$, the corresponding *lower expectation* of $X \in \mathbb{R}^\Omega$ is

$$\underline{\mathbb{E}}_{\mathcal{P}}[X] = \min_{P\in\mathcal{P}} \mathbb{E}_P[X]. \tag{3}$$

In general [37,41], we have that $\mathbb{C}_{\underline{P}}[X] \leq \underline{\mathbb{E}}_{\mathbf{core}(\underline{P})}[X] \leq \underline{\mathbb{E}}_{\mathcal{P}}[X]$, where the two inequalities can be strict. Nevertheless, in the particular case $\underline{P}$ is (at least) 2-monotone (see, e.g., [14,18]), then $\mathbb{C}_{\underline{P}}[X] = \underline{\mathbb{E}}_{\mathbf{core}(\underline{P})}[X]$ and $\underline{\mathbb{E}}_{\mathbf{core}(\underline{P})}[X] = \underline{\mathbb{E}}_{\mathcal{P}}[X]$ if $\mathcal{P}$ has the same extreme points of $\mathbf{core}(\underline{P})$.

3. DS-approximation of joint lower distributions with an independent ambiguous Bernoulli marginal

Let X, Y be discrete random variables taking values in $\mathcal{X} = \{x_1, \ldots, x_t\}$ and $\mathcal{Y} = \{0, 1\}$. Assume that no logical relations (structural zeros) are present between X and Y, therefore, we can simply identify X and Y with the projection maps on the product measurable space $(\mathcal{X} \times \mathcal{Y}, 2^{\mathcal{X}\times\mathcal{Y}})$. We also denote by $\widetilde{2^{\mathcal{X}}}$ and $\widetilde{2^{\mathcal{Y}}}$ the sub-algebras of $2^{\mathcal{X}\times\mathcal{Y}}$ isomorphic to $2^{\mathcal{X}}$ and $2^{\mathcal{Y}}$, respectively.

In what follows, we consider marginal *probability distributions* for X and Y, i.e., probability measures on the power sets of their ranges, respectively. To avoid cumbersome notation, we write $P_X(x) := P_X(\{x\})$ and $P_Y(y) := P_Y(\{y\})$, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

Let $P_X : 2^{\mathcal{X}} \to [0, 1]$ be a probability distribution for X and

$$\mathcal{P}_Y = \{P_Y^p : 2^{\mathcal{Y}} \to [0, 1] \,:\, p \in [p_1, p_2]\}$$

be a family of probability distributions for Y, where

$$P_Y^p(1) = p \quad \text{and} \quad P_Y^p(0) = 1 - p, \quad \text{with } 0 \leq p_1 \leq p_2 \leq 1.$$

Suppose that, for every $P_Y^p \in \mathcal{P}_Y$, the random variables X, Y are stochastically independent and the joint probability distribution $P^p : 2^{\mathcal{X}\times\mathcal{Y}} \to [0, 1]$ of the vector (X, Y) is obtained extending by additivity the assessment

$$P^p(\{(x, y)\}) = P_X(x) \cdot P_Y^p(y), \quad \text{for all } (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{4}$$

The above hypothesis of independence generalizes the original model given in [2], that is recovered for $p_1 = p_2$. Therefore, we get the family of joint distributions

$$\mathcal{P} = \{P^p \,:\, P^p \text{ is a joint distribution of } (X, Y) \text{ given by (4)}, p \in [p_1, p_2]\}. \tag{5}$$

Proposition 1. *The set $\mathcal{P}$ is a closed and convex subset of $[0, 1]^{2^{\mathcal{X}\times\mathcal{Y}}}$ endowed with the product topology, and its extreme points form the set $\mathbf{ext}(\mathcal{P}) = \{P^{p_1}, P^{p_2}\}$.*

Proof. The set $\mathcal{P}_Y$ of marginal probability distributions for Y is a closed and convex subset of $[0,1]^{2^{\mathcal{Y}}}$ endowed with the product topology, and $P^p(\mathcal{X}\times\{y\}) = P_Y^p(y)$, for all $p \in [p_1, p_2]$. In turn, this implies that every sequence $\{P^{p_n}\}_{n\in\mathbb{N}}$ in $\mathcal{P}$ converging pointwise on $2^{\mathcal{X}\times\mathcal{Y}}$ has a limit $P = P^p \in \mathcal{P}$, the convex combination of $P^p, P^{p'} \in \mathcal{P}$ with $\alpha \in [0,1]$ is such that $P = \alpha P^p + (1-\alpha)P^{p'} \in \mathcal{P}$, and $\mathbf{ext}(\mathcal{P}) = \{P^{p_1}, P^{p_2}\}$. $\square$

Let $\underline{P} = \min \mathcal{P}$ be the lower envelope of $\mathcal{P}$ and denote by $\underline{P}_Y = \min \mathcal{P}_Y$. The following example shows that $\underline{P}$ is generally not 2-monotone and we also have that $\mathcal{P}$ is strictly contained in $\mathbf{core}(\underline{P})$.

Example 1. For $\mathcal{X} = \{x_1, x_2\}$, denote

$$\mathcal{X}\times\mathcal{Y} = \{\underbrace{(x_1,1)}_{=a_1}, \underbrace{(x_1,0)}_{=a_2}, \underbrace{(x_2,1)}_{=a_3}, \underbrace{(x_2,0)}_{=a_4}\},$$

and let $A_i = \{a_i\}$, $A_{ij} = \{a_i, a_j\}$, $A_{ijk} = \{a_i, a_j, a_k\}$ and $A_{1234} = \mathcal{X}\times\mathcal{Y}$. Take the marginal probability distributions such that $P_X(x_1) = \frac{3}{4}$, $P_X(x_2) = \frac{1}{4}$, $P_Y^p(1) = p$, $P_Y^p(0) = 1-p$, where $p \in \left[\frac{1}{4}, \frac{3}{4}\right]$. The family $\mathcal{P}$ of joint probability distributions for (X,Y) has extreme points and lower envelope reported below:

$2^{\mathcal{X}\times\mathcal{Y}}$	$\emptyset$	A_1	A_2	A_3	A_4	A_{12}	A_{13}	A_{14}	A_{23}	A_{24}	A_{34}	A_{123}	A_{124}	A_{134}	A_{234}	A_{1234}
P^{p_1}	0	$\frac{3}{16}$	$\frac{9}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{12}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{10}{16}$	$\frac{12}{16}$	$\frac{4}{16}$	$\frac{13}{16}$	$\frac{15}{16}$	$\frac{7}{16}$	$\frac{13}{16}$	1
P^{p_2}	0	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{12}{16}$	$\frac{12}{16}$	$\frac{10}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{15}{16}$	$\frac{13}{16}$	$\frac{13}{16}$	$\frac{7}{16}$	1
$\underline{P}$	0	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{12}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{13}{16}$	$\frac{13}{16}$	$\frac{7}{16}$	$\frac{7}{16}$	1

The lower envelope is easily seen not to be 2-monotone since

$$\underline{P}(A_{123}) = \frac{13}{16} < \frac{15}{16} = \underline{P}(A_{12}) + \underline{P}(A_{23}) - \underline{P}(A_2).$$

We also have that $\mathcal{P} \subset \mathbf{core}(\underline{P})$ since $\mathbf{ext}(\mathbf{core}(\underline{P})) = \{P_1, P_2, P_3, P_4\}$, where, identifying each probability distribution on $2^{\mathcal{X}\times\mathcal{Y}}$ with the vector of its values on the atoms of $2^{\mathcal{X}\times\mathcal{Y}}$ we have

$$P_1 = P^{p_1} \equiv \left(\tfrac{3}{16}, \tfrac{9}{16}, \tfrac{1}{16}, \tfrac{3}{16}\right), \qquad P_2 = P^{p_2} \equiv \left(\tfrac{9}{16}, \tfrac{3}{16}, \tfrac{3}{16}, \tfrac{1}{16}\right),$$

$$P_3 \equiv \left(\tfrac{7}{16}, \tfrac{5}{16}, \tfrac{1}{16}, \tfrac{3}{16}\right), \qquad P_4 \equiv \left(\tfrac{5}{16}, \tfrac{7}{16}, \tfrac{3}{16}, \tfrac{1}{16}\right).$$

The inclusion $\mathcal{P} \subset \mathbf{core}(\underline{P})$ is due to fact that $\mathcal{P}$ is the credal set associated with a coherent lower prevision functional, while $\mathbf{core}(\underline{P})$ is the credal set induced by its restriction on indicators, that is $\underline{P}$ (see, e.g., [37,41]). In turn, this is another motivation to the failure of 2-monotonicity for $\underline{P}$: 2-monotonicity would imply $\mathcal{P} = \mathbf{core}(\underline{P})$. Actually, the failure of 2-monotonicity is not surprising since the set $\mathcal{P}$ can be considered as a particular strong product of two imprecise marginal probabilities [9,10].

Despite the failure of 2-monotonicity, the set $\mathcal{P}$ assures an important decomposition property, when computing lower expectations. The property reported in Proposition 2 recalls the *external additivity* property, discussed in the context of independent products [11,12,39].

Proposition 2. *For every $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ we have*

$$\underline{\mathbb{E}}_{\mathcal{P}}[f(X) + g(X,Y)] = \mathbb{E}_{P_X}[f(X)] + \underline{\mathbb{E}}_{\mathcal{P}}[g(X,Y)],$$

where $\mathbb{E}_{P_X}$ denotes the expectation with respect to the marginal P_X.

Proof. Every $P \in \mathcal{P}$ is such that $P_{|\widehat{2^{\mathcal{X}}}}$ coincides with the marginal distribution of X, therefore

$$\begin{aligned}\underline{\mathbb{E}}_{\mathcal{P}}[f(X) + g(X,Y)] &= \min_{P\in\mathcal{P}} \mathbb{E}_P[f(X) + g(X,Y)] \\ &= \min_{P\in\mathcal{P}} \left(\mathbb{E}_{P_X}[f(X)] + \mathbb{E}_P[g(X,Y)]\right) \\ &= \mathbb{E}_{P_X}[f(X)] + \min_{P\in\mathcal{P}} \mathbb{E}_P[g(X,Y)] \\ &= \mathbb{E}_{P_X}[f(X)] + \underline{\mathbb{E}}_{\mathcal{P}}[g(X,Y)]. \quad \square\end{aligned}$$

Due to the desirable property expressed after equation (3), we look for a belief function that approximates the lower probability $\underline{P}$ by following [21,22,25,26], so as to reduce lower expectations to Choquet expectations. We refer to it as *DS-approximation* (where "DS" stands for Dempster and Shafer). The decomposition property reported in Proposition 2 reveals to be of particular importance in the reinsurance application faced in Section 4, to compute reinsurance premia. For this reason, we seek DS-approximation schemes enforcing an analogous decomposition property with Choquet expectations. In turn, for this to hold, a necessary and sufficient condition is that the restriction of the DS-approximation of $\underline{P}$ on $2^{\widetilde{\mathcal{X}}}$ coincides with P_X, as this implies that all dominating joint probabilities have the same property.

3.1. Inner DS-approximations

We start our analysis by considering inner DS-approximations, for which the quoted decomposition property is automatically satisfied. Indeed, since $\underline{P}_{|2^{\widetilde{\mathcal{X}}}}$ coincides with P_X, every lower probability $\underline{Q}$ inner approximating $\underline{P}$, i.e., satisfying $\underline{Q} \geq \underline{P}$ pointwise on $2^{\mathcal{X}\times\mathcal{Y}}$, inherits the same property. We notice that in financial applications, inner DS-approximations are preferred, due to the phenomenon of dilation in lower-upper prices induced by other approximation schemes [8].

Searching for an *inner DS-approximation* (see [21]) means to look for a belief function Bel^i that dominates $\underline{P}$, i.e., $Bel^i \geq \underline{P}$ pointwise on $2^{\mathcal{X}\times\mathcal{Y}}$, and is as close as possible to $\underline{P}$ according to the squared Euclidean distance D_2 defined over the set of lower probabilities on $2^{\mathcal{X}\times\mathcal{Y}}$:

$$\text{minimize } D_2(\underline{P}, Bel)$$
$$\text{subject to:}$$
$$\begin{cases} \sum\limits_{B\subseteq A} m(B) \geq \underline{P}(A), & \text{for all } A \in 2^{\mathcal{X}\times\mathcal{Y}}, \\ \sum\limits_{B\subseteq \mathcal{X}\times\mathcal{Y}} m(B) = 1, & \\ m(B) \geq 0, & \text{for all } B \in 2^{\mathcal{X}\times\mathcal{Y}}, \\ m(\emptyset) = 0. & \end{cases} \tag{6}$$

We have that D_2, besides assuring uniqueness of the inner DS-approximation, has a justification in terms of a penalty coherence condition for belief functions [32]. Indeed, D_2 turns out to be the *Bregman divergence* [6] induced by the *Brier scoring rule* [5] that is used in [32] to provide an interpretation of belief assessments as forecasts. Thus, D_2 allows us to define the notion of *projection* onto a convex set of belief functions that, in this case, reduces to the classical orthogonal projection. The function D_2 is also referred to as *quadratic distance*, however, other distances for lower probabilities can be considered [4,21,22,25,26,33]. In particular, in [21] the authors consider inner approximations by means of D_2 together with some particular subfamilies of belief functions. We point out that, besides the quoted distances and divergences, several other choices are available, like minimizing the Kullback-Leibler divergence [27,28] or a measure of nonspecificity (or imprecision) [15].

It trivially holds that there are infinitely many inner DS-approximations of $\underline{P}$, as every P^p will work, so problem (6) is always feasible.

The following Example 2 shows that the D_2-optimal inner DS-approximation Bel^i of $\underline{P}$ is generally non-additive, though it may be additive for particular parameter settings. Nevertheless, the same example shows that, even though $\mathbf{core}(Bel^i) \subset \mathbf{core}(\underline{P})$, we have that $\mathbf{core}(Bel^i) \nsubseteq \mathcal{P}$ and $\mathcal{P} \nsubseteq \mathbf{core}(Bel^i)$. This last fact has important consequences when computing lower expectations, since no dominance relation can be established between the lower expectation computed with respect to $\mathcal{P}$ and the Choquet integral with respect to Bel^i.

Example 2. Let $\underline{P}$ be as in Example 1. The D_2-optimal inner DS-approximation Bel^i of $\underline{P}$ and its Möbius inverse m^i are reported below:

$2^{\mathcal{X}\times\mathcal{Y}}$	$\emptyset$	A_1	A_2	A_3	A_4	A_{12}	A_{13}	A_{14}	A_{23}	A_{24}	A_{34}	A_{123}	A_{124}	A_{134}	A_{234}	A_{1234}
m^i	0	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{4}{16}$	0	0	0	0	0	0	0	0	0	0
Bel^i	0	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{12}{16}$	$\frac{6}{16}$	$\frac{6}{16}$	$\frac{6}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{14}{16}$	$\frac{14}{16}$	$\frac{8}{16}$	$\frac{8}{16}$	1

We have that $\mathbf{ext}(\mathbf{core}(Bel^i)) = \{Q_1, Q_2\}$ where

$$Q_1 \equiv \left(\frac{4}{16}, \frac{8}{16}, \frac{2}{16}, \frac{2}{16}\right) \quad \text{and} \quad Q_2 \equiv \left(\frac{8}{16}, \frac{4}{16}, \frac{2}{16}, \frac{2}{16}\right).$$

Though $\mathbf{core}(Bel^i) \subset \mathbf{core}(\underline{P})$, since none between Q_1, Q_2 can be expressed as the convex combination of P^{p_1}, P^{p_2} and vice versa, it follows that between $\mathbf{core}(Bel^i)$ and $\mathcal{P}$ no containment relationship holds.

3.2. X-preserving outer DS-approximations

Example 2 suggests to move towards an outer DS-approximation. In this case, we search for a belief function Bel^o that is dominated by $\underline{P}$, i.e., $Bel^o \leq \underline{P}$ pointwise on $2^{\mathcal{X}\times\mathcal{Y}}$, and is as close as possible to $\underline{P}$ according to D_2.

As already noticed, the decomposition property expressed in Proposition 2, which is a desideratum in the reinsurance problem we face in Section 4, is implied by the fact that $\underline{P}_{|\widetilde{2^{\mathcal{X}}}}$ coincides with the probability distribution of X. This property is inherited by any inner DS-approximation Bel^i but generally not by an outer approximation Bel^o. Thus, we search for an outer approximation Bel^{oo} such that $Bel^{oo}_{|\widetilde{2^{\mathcal{X}}}}$ coincides with the marginal probability distribution of X. Such an outer DS-approximation will be called *X-preserving* and can be found solving the following optimization problem

$$\begin{aligned}&\text{minimize } D_2(\underline{P}, Bel)\\&\qquad\text{subject to:}\\&\begin{cases}\displaystyle\sum_{B\subseteq A} m(B) = \underline{P}(A), & \text{for all } A \in \widetilde{2^{\mathcal{X}}},\\ \displaystyle\sum_{B\subseteq A} m(B) \leq \underline{P}(A), & \text{for all } A \in 2^{\mathcal{X}\times\mathcal{Y}} \setminus \widetilde{2^{\mathcal{X}}},\\ \displaystyle\sum_{B\subseteq \mathcal{X}\times\mathcal{Y}} m(B) = 1, & \\ m(B) \geq 0, & \text{for all } B \in 2^{\mathcal{X}\times\mathcal{Y}},\\ m(\emptyset) = 0.\end{cases}\end{aligned} \tag{7}$$

Denoting by P_{X*} the inner measure induced by P_X on $2^{\mathcal{X}\times\mathcal{Y}}$, defined, for all $A \in 2^{\mathcal{X}\times\mathcal{Y}}$, as

$$P_{X*}(A) = \sup\left\{\sum_{x\in B} P_X(x) \,:\, B \times \mathcal{Y} \subseteq A,\, B \in 2^{\mathcal{X}}\right\},$$

P_{X*} is an X-preserving outer DS-approximation of $\underline{P}$. In turn, this implies that problem (7) is always feasible, and taking D_2 it admits a unique optimal solution.

Let us notice that the set of X-preserving outer DS-approximations of $\underline{P}$ is a convex subset of $[0,1]^{2^{\mathcal{X}\times\mathcal{Y}}}$, and Bel^{oo} turns out to be the orthogonal projection of $\underline{P}$ onto such set.

Example 3. Let $\underline{P}$ be as in Example 1. The D_2-optimal X-preserving outer DS-approximation Bel^{oo} of $\underline{P}$ and its Möbius inverse m^{oo} are reported below:

$2^{\mathcal{X}\times\mathcal{Y}}$	$\emptyset$	A_1	A_2	A_3	A_4	A_{12}	A_{13}	A_{14}	A_{23}	A_{24}	A_{34}	A_{123}	A_{124}	A_{134}	A_{234}	A_{1234}
m^{oo}	0	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{6}{16}$	0	0	0	0	$\frac{2}{16}$	0	0	0	0	0
Bel^{oo}	0	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{12}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{13}{16}$	$\frac{13}{16}$	$\frac{7}{16}$	$\frac{7}{16}$	1

We have that $\mathbf{ext}(\mathbf{core}(Bel^{oo})) = \{Q_1, Q_2, Q_3, Q_4\}$ where

$$Q_1 = P^{p_1} \equiv \left(\frac{3}{16}, \frac{9}{16}, \frac{1}{16}, \frac{3}{16}\right), \qquad Q_2 = P^{p_2} \equiv \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right),$$
$$Q_3 \equiv \left(\frac{3}{16}, \frac{9}{16}, \frac{3}{16}, \frac{1}{16}\right), \qquad Q_4 \equiv \left(\frac{9}{16}, \frac{3}{16}, \frac{1}{16}, \frac{3}{16}\right).$$

We stress that the first set of equalities appearing in system (7) guarantees that each belief function Bel we consider in the optimization problem satisfies $Bel_{|\widetilde{2^{\mathcal{X}}}} = P_X$. If such equalities are relaxed to less than or equal to inequalities, then the optimal solution turns out to be the D_2-optimal outer DS-approximation Bel^o. Though Bel^o is generally closer to $\underline{P}$ with respect to Bel^{oo} in terms of squared Euclidean distance, it generally fails the X-preserving property, as shown in the following example.

Example 4. Let $\underline{P}$ be as in Example 1 and Bel^{oo} as in Example 3. The D_2-optimal outer DS-approximation Bel^o of $\underline{P}$ and its Möbius inverse m^o are reported below:

$2^{\mathcal{X}\times\mathcal{Y}}$	$\emptyset$	A_1	A_2	A_3	A_4	A_{12}	A_{13}	A_{14}	A_{23}	A_{24}	A_{34}	A_{123}	A_{124}	A_{134}	A_{234}	A_{1234}
m^o	0	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{5}{16}$	0	$\frac{1}{16}$	$\frac{1}{16}$	0	$\frac{1}{16}$	0	0	0	0	0
Bel^o	0	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{11}{16}$	$\frac{4}{16}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{13}{16}$	$\frac{13}{16}$	$\frac{7}{16}$	$\frac{7}{16}$	1

It holds that $D_2(\underline{P}, Bel^o) = \frac{4}{256} < \frac{8}{256} = D_2(\underline{P}, Bel^{oo})$, nevertheless, it is easy to verify that $Bel^o_{|\widetilde{2^{\mathcal{X}}}}$ does not coincide with P_X.

The following theorem provides a characterization of the D_2-optimal X-preserving outer DS-approximation of $\underline{P}$.

Theorem 1. *The D_2-optimal X-preserving outer DS-approximation Bel^{oo} of $\underline{P}$ has Möbius inverse $m^{oo} : 2^{\mathcal{X}\times\mathcal{Y}} \to [0,1]$ satisfying the following properties:*

(i) $m^{oo}(\{(x,y)\}) = P_X(x)\underline{P}_Y(y)$, *for all* $(x,y) \in \mathcal{X}\times\mathcal{Y}$,
(ii) $m^{oo}(\{x\}\times\mathcal{Y}) = P_X(x) - \sum_{y\in\mathcal{Y}} P_X(x)\underline{P}_Y(y)$, *for all* $x \in \mathcal{X}$,
(iii) m^{oo} *is* 0 *on* $2^{\mathcal{X}\times\mathcal{Y}} \setminus (\{\{(x,y)\} : (x,y)\in\mathcal{X}\times\mathcal{Y}\} \cup \{\{x\}\times\mathcal{Y} : x\in\mathcal{X}\})$.

Proof. The proof consists in proving the following points:

1. m^{oo} satisfying *(i)–(iii)* is the Möbius inverse of an X-preserving outer DS-approximation Bel^{oo} of $\underline{P}$;
2. every X-preserving outer DS-approximation Bel of $\underline{P}$ with Möbius inverse m satisfies the properties
   (a) $Bel(\{(x,y)\}) \leq Bel^{oo}(\{(x,y)\})$, for all $(x,y)\in\mathcal{X}\times\mathcal{Y}$;
   (b) $\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} m(\{(x,y)\}) + \sum_{x\in\mathcal{X}} m(\{x\}\times\mathcal{Y}) = 1$;
3. if an X-preserving outer DS-approximation Bel of $\underline{P}$ is such that $Bel(\{(x,y)\}) = Bel^{oo}(\{(x,y)\})$, for all $(x,y)\in\mathcal{X}\times\mathcal{Y}$, then $Bel = Bel^{oo}$;
4. Bel^{oo} minimizes D_2 in the class of X-preserving outer DS-approximations of $\underline{P}$.

Point 1. For all $(x,y)\in\mathcal{X}\times\mathcal{Y}$, we have that

$$\underline{P}(\{(x,y)\}) = P_X(x)\underline{P}_Y(y) = m^{oo}(\{(x,y)\}) = Bel^{oo}(\{(x,y)\}),$$

and, for all $x\in\mathcal{X}$, we have that

$$\underline{P}(\{x\}\times\mathcal{Y}) = P_X(x) = m^{oo}(\{x\}\times\mathcal{Y}) + \sum_{y\in\mathcal{Y}} m^{oo}(\{(x,y)\}) = Bel^{oo}(\{x\}\times\mathcal{Y}),$$

which implies that, for all $A\in 2^{\mathcal{X}}$, it holds that

$$\underline{P}(A\times\mathcal{Y}) = \sum_{x\in A} P_X(x) = Bel^{oo}(A\times\mathcal{Y}).$$

More generally, for all $B\in 2^{\mathcal{X}\times\mathcal{Y}}$, we have that

$$\begin{aligned}
\underline{P}(B) &= \sum_{\{x\}\times\mathcal{Y}\subseteq B} P_X(x) + \min_{i=1,2}\left\{\sum_{\substack{(x,y)\in B\\ \{x\}\times\mathcal{Y}\not\subseteq B}} P_X(x)P_Y^{p_i}(y)\right\}\\
&\geq \sum_{\{x\}\times\mathcal{Y}\subseteq B} P_X(x) + \sum_{\substack{(x,y)\in B\\ \{x\}\times\mathcal{Y}\not\subseteq B}} P_X(x)\underline{P}_Y(y)\\
&= \sum_{C\subseteq B} m^{oo}(C) = Bel^{oo}(B),
\end{aligned}$$

therefore, Bel^{oo} is an X-preserving outer DS-approximation of $\underline{P}$.

Point 2. By point 1 it immediately follows that any X-preserving outer DS-approximation Bel of $\underline{P}$ satisfies property *(a)* since, for all $(x,y)\in\mathcal{X}\times\mathcal{Y}$,

$$Bel(\{(x,y)\}) \leq \underline{P}(\{(x,y)\}) = Bel^{oo}(\{(x,y)\}).$$

On the other hand, since $\mathcal{Y} = \{0,1\}$, property *(b)* follows since

$$\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} m(\{(x,y)\}) + \sum_{x\in\mathcal{X}} m(\{x\}\times\mathcal{Y}) = \sum_{x\in\mathcal{X}} Bel(\{x\}\times\mathcal{Y}) = \sum_{x\in\mathcal{X}} P_X(x) = 1.$$

Point 3. Suppose Bel is an X-preserving outer DS-approximation of $\underline{P}$ such that $Bel(\{(x,y)\}) = Bel^{oo}(\{(x,y)\})$, for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$, and let m be its Möbius inverse. We have that, for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$Bel(\{(x,y)\}) = m(\{(x,y)\}) = m^{oo}(\{(x,y)\}) = Bel^{oo}(\{(x,y)\}),$$

moreover, since Bel is X-preserving, for all $x \in \mathcal{X}$, it must hold

$$\begin{aligned} m(\{x\} \times \mathcal{Y}) &= P_X(x) - \sum_{y \in \mathcal{Y}} m(\{(x,y)\}) \\ &= P_X(x) - \sum_{y \in \mathcal{Y}} m^{oo}(\{(x,y)\}) = m^{oo}(\{x\} \times \mathcal{Y}), \end{aligned}$$

and since by property *(b)* of point 2

$$\begin{aligned} &\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} m(\{(x,y)\}) + \sum_{x \in \mathcal{X}} m(\{x\} \times \mathcal{Y}) \\ &= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} m^{oo}(\{(x,y)\}) + \sum_{x \in \mathcal{X}} m^{oo}(\{x\} \times \mathcal{Y}) = 1, \end{aligned}$$

it must be $m = m^{oo}$, which implies $Bel = Bel^{oo}$.

Point 4. Let $Bel \neq Bel^{oo}$ be an X-preserving outer DS-approximation of $\underline{P}$, having Möbius inverse m. By property *(a)* of point 2 and point 3, it must be $Bel(\{(x,y)\}) < Bel^{oo}(\{(x,y)\})$ for at least a $(x,y) \in \mathcal{X} \times \mathcal{Y}$, that is $m(\{(x,y)\}) < m^{oo}(\{(x,y)\})$. Since Bel satisfies property *(b)* of point 2 we get that m is 0 on $2^{\mathcal{X} \times \mathcal{Y}} \setminus (\{\{(x,y)\} : (x,y) \in \mathcal{X} \times \mathcal{Y}\} \cup \{\{x\} \times \mathcal{Y} : x \in \mathcal{X}\})$. From this we obtain that

$$D_2(\underline{P}, Bel)^2 = \sum_{B \in 2^{\mathcal{X} \times \mathcal{Y}}} \left(\underline{P}(B) - \sum_{\{x\} \times \mathcal{Y} \subseteq B} P_X(x) - \sum_{\substack{(x,y) \in B \\ \{x\} \times \mathcal{Y} \not\subseteq B}} m(\{(x,y)\}) \right)^2,$$

$$D_2(\underline{P}, Bel^{oo})^2 = \sum_{B \in 2^{\mathcal{X} \times \mathcal{Y}}} \left(\underline{P}(B) - \sum_{\{x\} \times \mathcal{Y} \subseteq B} P_X(x) - \sum_{\substack{(x,y) \in B \\ \{x\} \times \mathcal{Y} \not\subseteq B}} m^{oo}(\{(x,y)\}) \right)^2,$$

which implies that $D_2(\underline{P}, Bel^{oo}) < D_2(\underline{P}, Bel)$. □

The D_2-optimal X-preserving outer approximation Bel^{oo} is a joint belief function on $2^{\mathcal{X} \times \mathcal{Y}}$ that extends the precise marginal of X and the imprecise marginal of Y, that is P_X and $\underline{P}_Y$. Indeed, P_X and $\underline{P}_Y$ induce an additive belief function on $\widetilde{2^{\mathcal{X}}}$ and a non-additive belief function on $\widetilde{2^{\mathcal{Y}}}$, respectively, that are restrictions of Bel^{oo}.

We notice that, since $\mathcal{P} \subset \textbf{core}(Bel^{oo})$, then considering Bel^{oo} we are actually weakening the independence hypothesis between X and Y, that is $\textbf{core}(Bel^{oo})$ contains joint probability distributions on $2^{\mathcal{X} \times \mathcal{Y}}$ with marginals P_X and $P_Y^p \in \mathcal{P}_Y$, obtained using copulas (see [29]) different from the independence copula.

3.3. Properties of inner and X-preserving outer DS-approximations

Both Bel^i and Bel^{oo} allow the following decomposition of the corresponding Choquet expectation that is analogous to Proposition 2.

Proposition 3. *For every $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ we have:*

(i) $\mathbb{C}_{Bel^i}[f(X) + g(X,Y)] = \mathbb{E}_{P_X}[f(X)] + \mathbb{C}_{Bel^i}[g(X,Y)]$;
(ii) $\mathbb{C}_{Bel^{oo}}[f(X) + g(X,Y)] = \mathbb{E}_{P_X}[f(X)] + \mathbb{C}_{Bel^{oo}}[g(X,Y)]$;

where $\mathbb{E}_{P_X}$ denotes the expectation with respect to the marginal P_X.

Proof. We only prove *(i)* since the proof of *(ii)* is analogous. Every $P \in \textbf{core}(Bel^i)$ is such that $P_{|\widetilde{2^{\mathcal{X}}}}$ coincides with the marginal distribution of X, therefore

$$\begin{aligned} \mathbb{C}_{Bel^i}[f(X) + g(X,Y)] &= \min_{P \in \textbf{core}(Bel^i)} \mathbb{E}_P[f(X) + g(X,Y)] \\ &= \min_{P \in \textbf{core}(Bel^i)} \left(\mathbb{E}_{P_X}[f(X)] + \mathbb{E}_P[g(X,Y)] \right) \end{aligned}$$

$$= \mathbb{E}_{P_X}[f(X)] + \min_{P \in \mathbf{core}(Bel^i)} \mathbb{E}_P[g(X,Y)]$$
$$= \mathbb{E}_{P_X}[f(X)] + \mathbb{C}_{Bel^i}[g(X,Y)]. \quad \square$$

As recalled in Section 2, for representing a Choquet expectation as a lower expectation with respect to the core, 2-monotonicity is sufficient. Therefore, an interesting problem is to look for a D_2-optimal inner [X-preserving outer] approximation in the space of 2-monotone lower probabilities on $2^{\mathcal{X}\times\mathcal{Y}}$, namely a D_2-optimal *inner [X-preserving outer] 2M-approximation* of $\underline{P}$, denoted as $\underline{Q}^i$ [$\underline{Q}^{oo}$]. Notice that the problem is well-posed since inner [X-preserving outer] DS-approximations are particular inner [X-preserving outer] 2M-approximations. The following theorem proves that every inner [X-preserving outer] 2M-approximation is an inner [X-preserving outer] DS-approximation, therefore $\underline{Q}^i = Bel^i$ [$\underline{Q}^{oo} = Bel^{oo}$]. In turn, this implies that working in the framework of belief functions is not a real restriction in this particular problem.

Theorem 2. *The following statements hold:*

(i) every inner 2M-approximation of $\underline{P}$ is an inner DS-approximation of $\underline{P}$.
(ii) every X-preserving outer 2M-approximation of $\underline{P}$ is an X-preserving outer DS-approximation of $\underline{P}$.

Proof. Define the set

$$\mathcal{V} = 2^{\mathcal{X}\times\mathcal{Y}} \setminus (\{\{(x,y)\} : (x,y) \in \mathcal{X}\times\mathcal{Y}\} \cup \{\{x\}\times\mathcal{Y} : x \in \mathcal{X}\} \cup \{\emptyset\}).$$

Statement (i). Let $\underline{Q}$ be an inner 2M-approximation of $\underline{P}$ with Möbius inverse m. By equations (1) and (2), since $\underline{Q}_{|\widetilde{2^{\mathcal{X}}}}$ coincides with P_X and $\underline{P}(\{(x,y)\}) = P_X(x)\underline{P}_Y(y)$, for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$, we get that:

- $m(\emptyset) = 0$;
- $P_X(x)\underline{P}_Y(y) \le m(\{(x,y)\}) \le P_X(x)$, for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$;
- $m(\{x\}\times\mathcal{Y}) = P_X(x) - \sum_{y\in\mathcal{Y}} m(\{(x,y)\}) \ge 0$, for all $x \in \mathcal{X}$;
- $\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} m(\{(x,y)\}) + \sum_{x\in\mathcal{X}} m(\{x\}\times\mathcal{Y}) = \sum_{x\in\mathcal{X}} P_X(x) = 1$;
- $\sum_{B\in\mathcal{V}} m(B) = 0$.

We show that $m(B) = 0$, for all $B \in \mathcal{V}$. The set $\mathcal{V}$ can be partitioned, generally not in a unique way, in a finite number of disjoint sets $\mathcal{V}_1, \ldots, \mathcal{V}_k$, where each $\mathcal{V}_h$ contains a maximal element $A_h \in 2^{\mathcal{X}\times\mathcal{Y}}$ and a minimal element $\{(x_i, y_{i'}), (x_j, y_{j'})\} \in 2^{\mathcal{X}\times\mathcal{Y}}$, $(x_i, y_{i'}) \neq (x_j, y_{j'})$, such that

$$\mathcal{V}_h = \{B \in \mathcal{V} : \{(x_i, y_{i'}), (x_j, y_{j'})\} \subseteq B \subseteq A_h\}. \tag{8}$$

By equation (2), for $h = 1, \ldots, k$, it holds that

$$\sum_{B\in\mathcal{V}_h} m(B) \ge 0,$$

and since

$$\sum_{B\in\mathcal{V}} m(B) = \sum_{h=1}^{k} \sum_{B\in\mathcal{V}_h} m(B) = 0,$$

we get that, for $h = 1, \ldots, k$, it holds that

$$\sum_{B\in\mathcal{V}_h} m(B) = 0.$$

Finally, varying all the possible partitions of $\mathcal{V}$ in sets of the type (8) we derive that $m(B) = 0$, for all $B \in \mathcal{V}$. Hence, since m ranges in $[0,1]$, we have that $\underline{Q}$ is a belief function, therefore it is an inner DS-approximation of $\underline{P}$.

Statement (ii). Let $\underline{Q}$ be an X-preserving outer 2M-approximation of $\underline{P}$ with Möbius inverse m. By equations (1) and (2), since $\underline{Q}_{|\widetilde{2^{\mathcal{X}}}}$ coincides with P_X and $\underline{P}(\{(x,y)\}) = P_X(x)\underline{P}_Y(y)$, for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$, we get that:

- $m(\emptyset) = 0$;
- $0 \le m(\{(x,y)\}) \le P_X(x)\underline{P}_Y(y)$, for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$;

- $m(\{x\} \times \mathcal{Y}) = P_X(x) - \sum_{y \in \mathcal{Y}} m(\{(x,y)\}) \geq 0$, for all $x \in \mathcal{X}$;
- $\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} m(\{(x,y)\}) + \sum_{x \in \mathcal{X}} m(\{x\} \times \mathcal{Y}) = \sum_{x \in \mathcal{X}} P_X(x) = 1$;
- $\sum_{B \in \mathcal{V}} m(B) = 0$.

Proceeding as in the proof of statement *(i)*, we show that $m(B) = 0$, for all $B \in \mathcal{V}$. Hence, since m ranges in $[0,1]$, we have that $\underline{Q}$ is a belief function, therefore it is an X-preserving outer DS-approximation of $\underline{P}$. $\square$

4. Ambiguous randomized reinsurance stop-loss treaties

Referring to X, Y of Section 3, here the variable X denotes the (non-negative) aggregate loss of an insurer over one year, while Y (indicating reinsurance) is an ambiguous Bernoulli random variable independent of X, under the marginal probability distribution P_X and any marginal probability distribution in $\mathcal{P}_Y$. In this setting, we assume that $x_1 = \min \mathcal{X} = 0$, that stands for a null aggregate loss, moreover, we suppose $P_X(x) > 0$, for all $x \in \mathcal{X}$, and denote again $\underline{P}_Y = \min \mathcal{P}_Y$.

We further denote by F_X the cumulative distribution function of X induced by P_X, defined, for every $x \geq 0$, as

$$F_X(x) = \sum_{x_j \leq x} P_X(x_j).$$

In what follows, we refer to the set $\mathcal{P}$ of joint distributions of (X, Y) given by (5) and let $\underline{P} = \min \mathcal{P}$.

Following [2], we consider a reinsurance contract in which the retained loss of the insurer is singled out by the random variable

$$r(X,Y,d) = \begin{cases} \min(X,d) & \text{if } Y = 1, \\ X & \text{if } Y = 0, \end{cases} \tag{9}$$

where $d \geq 0$ denotes the retention in a stop-loss treaty. For every $P^p \in \mathcal{P}$

$$\mathbb{E}_{P^{p_2}}[r(X,Y,d)] \leq \mathbb{E}_{P^p}[r(X,Y,d)] \leq \mathbb{E}_{P^{p_1}}[r(X,Y,d)].$$

Let $\pi(X)$ and $\pi_R(d)$ be, respectively, the total premium the insurer receives from the policyholders for the aggregate loss X and the premium required from the reinsurer for a randomized stop-loss treaty with retention level d. By adopting the *expected value principle* (see [1]) with *safety loading* $\theta > 0$, and assuming a pessimistic attitude towards ambiguity, by virtue of Proposition 2, we set

$$\begin{aligned} \pi_R(d) &= (1+\theta)\underline{\mathbb{E}}_{\mathcal{P}}[X - r(X,Y,d)] \\ &= (1+\theta)\left(\mathbb{E}_{P_X}[X] + \underline{\mathbb{E}}_{\mathcal{P}}[-r(X,Y,d)]\right). \end{aligned} \tag{10}$$

We introduce an ambiguous version of *value-at-risk* (see, e.g., [19]) as risk measure associated with $r(X,Y,d)$, defined, for $\alpha \in (0,1)$, as

$$\underline{\mathrm{VaR}}_\alpha(r(X,Y,d)) := \inf\{x : \underline{F}_{r(X,Y,d)}(x) \geq \alpha\}, \tag{11}$$

where $\underline{F}_{r(X,Y,d)}(x) := \underline{P}(r(X,Y,d) \leq x)$ is the lower cumulative distribution function of $r(X,Y,d)$ under $\underline{P}$. This particular choice for the solvency risk measure $\rho(r(X,Y,d))$ is in line with [2] and deals with ambiguity in a cautious way, since it refers to a lower cumulative distribution function.

The insurer's annual profit under reinsurance is

$$Z(X,Y,d) = \frac{\pi(X) - \pi_R(d)}{1 - r_{coc}} - r(X,Y,d) - \frac{r_{coc}}{1 - r_{coc}}\underline{\mathrm{VaR}}_\alpha(r(X,Y,d)), \tag{12}$$

where r_{coc} denotes the cost of capital rate. Due to the translation invariance of the lower expectation operator, we get that

$$\begin{aligned} \underline{\mathbb{E}}_{\mathcal{P}}[Z(X,Y,d)] = \frac{\pi(X) - \pi_R(d)}{1 - r_{coc}} &+ \underline{\mathbb{E}}_{\mathcal{P}}[-r(X,Y,d)] \\ &- \frac{r_{coc}}{1 - r_{coc}}\underline{\mathrm{VaR}}_\alpha(r(X,Y,d)). \end{aligned} \tag{13}$$

Under this pessimistic attitude towards ambiguity, the issue is to maximize $\underline{\mathbb{E}}_{\mathcal{P}}[Z(X,Y,d)]$ seen as a function of d.

Remark 1. The lower cumulative distribution function $\underline{F}_{r(X,Y,d)}(x)$ can be associated with an upper cumulative distribution function, defined as $\overline{F}_{r(X,Y,d)}(x) := 1 - \underline{P}(r(X,Y,d) > x)$. It is easily seen that $\underline{F}_{r(X,Y,d)}(x) \leq \overline{F}_{r(X,Y,d)}(x)$, thus the pair of functions $(\underline{F}_{r(X,Y,d)}, \overline{F}_{r(X,Y,d)})$ gives rise to a *p-box* [17,36]. Referring to the upper expectation functional $\overline{\mathbb{E}}_{\mathcal{P}}$ induced by $\mathcal{P}$ and to the upper cumulative distribution function $\overline{F}_{r(X,Y,d)}(x)$, an optimistic formulation of the optimization problem can be provided by suitably changing equations (10), (11), (12) and (13), and replacing the maximin criterion with the maximax criterion. What is important to notice is that, in doing so, the modified insurer's annual profit under reinsurance in (12) becomes a different random variable, due to the modification of (10) and (11). Thus, the optimistic version of (13) under $\overline{\mathbb{E}}_{\mathcal{P}}$ is actually a distinct optimization problem (duality does not hold, in general).

Example 5. Take X ranging in $\mathcal{X} = \{0, 100, 1000\}$ with probability distribution such that $P_X(0) = \frac{9}{10}, P_X(100) = \frac{6}{100}, P_X(1000) = \frac{4}{100}$, and let Y be an ambiguous Bernoulli random variable with probability distribution such that $P_Y^p(1) = p$, $P_Y^p(0) = 1 - p$ and $p \in \left[\frac{8}{10}, \frac{9}{10}\right]$. Let $\mathcal{P}$ be defined as in (5) and $\underline{P} = \min \mathcal{P}$ pointwise on $2^{\mathcal{X}\times\mathcal{Y}}$. Take $\alpha = 0.99$, $\theta = 0.1$, $r_{coc} = 0.07$, $\pi(X) = (1 + 0.1)\mathbb{E}_{P_X}[X] = 50.6$. Let $r(X, Y, d)$ be defined as in (9):

$$\underline{\mathbb{E}}_{\mathcal{P}}[-r(X,Y,d)] = \begin{cases} -0.08d - 9.2 & \text{if } 0 \le d < 100, \\ -0.032d - 14 & \text{if } 100 \le d < 1000, \\ -46 & \text{if } d \ge 1000, \end{cases}$$

therefore, we get

$$\pi_R(d) = \begin{cases} -0.088d + 40.48 & \text{if } 0 \le d < 100, \\ -0.0352d + 35.2 & \text{if } 100 \le d < 1000, \\ 0 & \text{if } d \ge 1000. \end{cases}$$

Moreover, $\underline{F}_{r(X,Y,d)}(x)$ has the following definitions, according to the value of d

$d = 0$

$$\underline{F}_{r(X,Y,d)}(x) = \begin{cases} 0 & \text{if } x < 0, \\ 0.98 & \text{if } 0 \le x < 100, \\ 0.992 & \text{if } 100 \le x < 1000, \\ 1 & \text{if } x \ge 1000, \end{cases}$$

$0 < d < 100$

$$\underline{F}_{r(X,Y,d)}(x) = \begin{cases} 0 & \text{if } x < 0, \\ 0.9 & \text{if } 0 \le x < d, \\ 0.98 & \text{if } d \le x < 100, \\ 0.992 & \text{if } 100 \le x < 1000, \\ 1 & \text{if } x \ge 1000, \end{cases}$$

$d = 100$

$$\underline{F}_{r(X,Y,d)}(x) = \begin{cases} 0 & \text{if } x < 0, \\ 0.9 & \text{if } 0 \le x < 100, \\ 0.992 & \text{if } 100 \le x < 100, \\ 1 & \text{if } x \ge 1000, \end{cases}$$

$100 < d < 1000$

$$\underline{F}_{r(X,Y,d)}(x) = \begin{cases} 0 & \text{if } x < 0, \\ 0.9 & \text{if } 0 \le x < 100, \\ 0.96 & \text{if } 100 \le x < d, \\ 0.992 & \text{if } d \le x < 1000, \\ 1 & \text{if } x \ge 1000, \end{cases}$$

$d \ge 1000$

$$\underline{F}_{r(X,Y,d)}(x) = \begin{cases} 0 & \text{if } x < 0, \\ 0.9 & \text{if } 0 \le x < 100, \\ 0.96 & \text{if } 100 \le x < 1000, \\ 1 & \text{if } x \ge 1000. \end{cases}$$

Fig. 1 shows the graph of $\underline{F}_{r(X,Y,d)}(x)$ for different values of d.

Referring to the different definitions of $\underline{F}_{r(X,Y,d)}(x)$, according to the value of d, we have that

$$\underline{\text{VaR}}_{\alpha}(r(X,Y,d)) = \begin{cases} 100 & \text{if } 0 \le d < 100, \\ d & \text{if } 100 \le d < 1000, \\ 1000 & \text{if } d \ge 1000. \end{cases}$$

Thus we get that

$$\underline{\mathbb{E}}_{\mathcal{P}}[Z(X,Y,d)] = \begin{cases} \left(\frac{0.088}{0.93} - 0.08\right)d + \left(\frac{3.12}{0.93} - 9.2\right) & \text{if } 0 \le d < 100, \\ \left(-\frac{0.0348}{0.93} - 0.032\right)d + \left(\frac{15.4}{0.93} - 14\right) & \text{if } 100 \le d < 1000, \\ -\frac{19.4}{0.93} - 46 & \text{if } d \ge 1000. \end{cases}$$

It is immediate to verify that $\underline{\mathbb{E}}_{\mathcal{P}}[Z(X,Y,d)]$, seen as a function of d, has a global maximum at $d^* = 100$.

In the precise case, that is taking $p_1 = p_2 = p$, the optimal retention level d^* is 100 for $p \in [0.8, 0.9)$ and 0 for $p = 0.9$. Thus, the optimal retention level in the imprecise case is greater than or equal to that in the precise case.

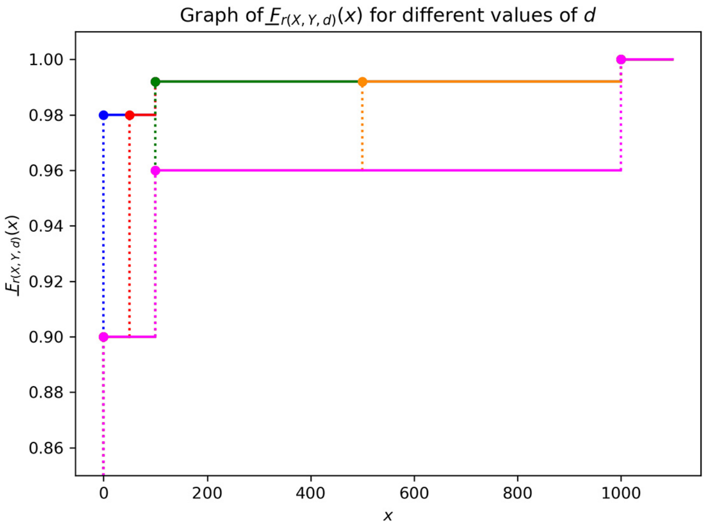

Fig. 1. Graph of $\underline{F}_{r(X,Y,d)}(x)$: $d=0$ in blue; $0<d<100$ in red; $d=100$ in green; $100<d<1000$ in orange; $d\geq 1000$ in magenta. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Despite using $\mathcal{P}$ and the associated lower expectation functional $\underline{\mathbb{E}}_{\mathcal{P}}$, we can consider the D_2-optimal X-preserving outer DS-approximation Bel^{oo} of $\underline{P}$ together with the corresponding Choquet expectation functional $\mathbb{C}_{Bel^{oo}}$. The use of Bel^{oo} and $\mathbb{C}_{Bel^{oo}}$ in place of $\mathcal{P}$ and $\underline{\mathbb{E}}_{\mathcal{P}}$, is motivated by the good computational properties of the Choquet expectation recalled in Section 2.

By virtue of Proposition 3 the premium is

$$\begin{aligned}\pi_R^{oo}(d) &= (1+\theta)\mathbb{C}_{Bel^{oo}}[X-r(X,Y,d)]\\ &= (1+\theta)\left(\mathbb{E}_{P_X}[X]+\mathbb{C}_{Bel^{oo}}[-r(X,Y,d)]\right),\end{aligned} \tag{14}$$

and the risk measure becomes

$$\underline{\mathrm{VaR}}_\alpha^{oo}(r(X,Y,d)) := \inf\{x : \underline{F}^{oo}_{r(X,Y,d)}(x)\geq\alpha\}, \tag{15}$$

where $\underline{F}^{oo}_{r(X,Y,d)}(x) := Bel^{oo}(r(X,Y,d)\leq x)$. The choice of this ambiguous version of value-at-risk, computed with respect to the lower cumulative distribution function $\underline{F}^{oo}_{r(X,Y,d)}(x)$, serves to favor a comparison with the original model that refers to $\mathcal{P}$.

The insurer's annual profit under reinsurance is then changed in

$$Z^{oo}(X,Y,d) = \frac{\pi(X)-\pi_R^{oo}(d)}{1-r_{coc}} - r(X,Y,d) - \frac{r_{coc}}{1-r_{coc}}\underline{\mathrm{VaR}}_\alpha^{oo}(r(X,Y,d)),$$

thus we get

$$\begin{aligned}\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)] &= \frac{\pi(X)-\pi_R^{oo}(d)}{1-r_{coc}} + \mathbb{C}_{Bel^{oo}}[-r(X,Y,d)]\\ &\quad - \frac{r_{coc}}{1-r_{coc}}\underline{\mathrm{VaR}}_\alpha^{oo}(r(X,Y,d)).\end{aligned} \tag{16}$$

The issue is to maximize $\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)]$ seen as a function of d. Analogously, we define $\pi_R^i(d)$, $\underline{\mathrm{VaR}}_\alpha^i$ and $Z^i(X,Y,d)$ when we use Bel^i and $\mathbb{C}_{Bel^i}$.

The following theorem shows that $\mathbb{C}_{Bel^{oo}}[-r(X,Y,d)]$ actually coincides with $\underline{\mathbb{E}}_{\mathcal{P}}[-r(X,Y,d)]$.

Theorem 3. *Let $\mathcal{X}=\{x_1,\ldots,x_t\}$, $d\geq 0$ and denote by $k(d)$ the maximum index in $\{1,\ldots,t\}$ such that $x_{k(d)}\leq d$. The following statement holds:*

$$\begin{aligned}\mathbb{C}_{Bel^{oo}}[-r(X,Y,d)] &= \underline{\mathbb{E}}_{\mathcal{P}}[-r(X,Y,d)]\\ &= -\mathbb{E}_{P_X}[X] + \sum_{j>k(d)}(x_j-d)P_X(x_j)\underline{P}_Y(1),\\ &= -\mathbb{E}_{P_X}[X] + (1-F_X(x_{k(d)}))\mathbb{E}_{P_X}[X-d|X>x_{k(d)}]\underline{P}_Y(1),\end{aligned}$$

where, in the last two equalities, the second term of the sum vanishes if $k(d) = t$.

Proof. Let m^{oo} be the Möbius inverse of Bel^{oo}. By Theorem 1, we have that

$$\begin{aligned}\mathbb{C}_{Bel^{oo}}[-r(X,Y,d)] &= -\sum_{j=1}^{k(d)} x_j P_X(x_j) - \sum_{j>k(d)} d \cdot m^{oo}(\{(x_j,1)\}) \\ &\quad - \sum_{j>k(d)} x_j(m^{oo}(\{(x_j,0)\}) + m^{oo}(\{x_j\}\times\mathcal{Y})) \\ &= -\mathbb{E}_{P_X}[X] + \sum_{j>k(d)} (x_j - d) m^{oo}(\{(x_j,1)\}) \\ &= -\mathbb{E}_{P_X}[X] + \sum_{j>k(d)} (x_j - d) P_X(x_j) \underline{P}_Y(1).\end{aligned}$$

On the other hand, it holds that

$$\begin{aligned}\underline{\mathbb{E}}_{\mathcal{P}}[-r(X,Y,d)] &= \min_{i=1,2}\left\{-\sum_{j=1}^{k(d)} x_j P_X(x_j) - \sum_{j>k(d)} d \cdot P_X(x_j)\cdot P_Y^{p_i}(1)\right. \\ &\quad \left. - \sum_{j>k(d)} x_j \cdot P_X(x_j)\cdot P_Y^{p_i}(1)\right\} \\ &= -\mathbb{E}_{P_X}[X] + \min_{i=1,2}\left\{\sum_{j>k(d)} (x_j - d)\cdot P_X(x_j)\cdot P_Y^{p_i}(1)\right\} \\ &= -\mathbb{E}_{P_X}[X] + \sum_{j>k(d)} (x_j - d) P_X(x_j)\underline{P}_Y(1),\end{aligned}$$

therefore, the statement follows since $\sum_{j>k(d)} (x_j - d) P_X(x_j)\underline{P}_Y(1)$ can be written as

$$(1 - F_X(x_{k(d)}))\mathbb{E}_{P_X}[X - d | X > x_{k(d)}]\underline{P}_Y(1). \quad \square$$

The following theorem shows that also $\underline{F}^{oo}_{r(X,Y,d)}(x)$ and $\underline{F}_{r(X,Y,d)}(x)$ coincide, i.e., the outer approximation on the left tails coincides with the original lower probability.

Theorem 4. *The following statement holds:*

$$\begin{aligned}\underline{F}^{oo}_{r(X,Y,d)}(x) &= \underline{F}_{r(X,Y,d)}(x) \\ &= \begin{cases} \sum_{x_j \le x} P_X(x_j) & \text{if } 0 \le x < d, \\ \sum_{x_j \le x} P_X(x_j) + \sum_{x_j > x} P_X(x_j)\underline{P}_Y(1) & \text{if } x \ge d, \end{cases} \\ &= \begin{cases} F_X(x) & \text{if } 0 \le x < d, \\ F_X(x) + (1 - F_X(x))\underline{P}_Y(1) & \text{if } x \ge d, \end{cases}\end{aligned}$$

where, in the last two equalities, the first branch vanishes if $d = 0$.

Proof. For $0 \le x < d$, it holds that

$$\{r(X,Y,d) \le x\} = \bigcup_{x_j \le x} (\{x_j\}\times\mathcal{Y}),$$

therefore, since both Bel^{oo} and $\underline{P}$ preserve the marginal on X, we get

$$Bel^{oo}(r(X,Y,d) \le x) = \underline{P}(r(X,Y,d) \le x) = \sum_{x_j \le x} P_X(x_j) = F_X(x).$$

For $x > d$, it holds that

$$\{r(X,Y,d) \leq x\} = \bigcup_{x_j \leq x} (\{x_j\} \times \mathcal{Y}) \cup \bigcup_{x_j > x} (\{x_j\} \times \{1\}).$$

Let m^{oo} be the Möbius inverse of Bel^{oo}. By Theorem 1 we get that

$$\begin{aligned} Bel^{oo}(r(X,Y,d) \leq x) &= \sum_{x_j \leq x} P_X(x_j) + \sum_{x_j > x} m^{oo}(\{x_j\} \times \{1\}) \\ &= \sum_{x_j \leq x} P_X(x_j) + \sum_{x_j > x} P_X(x_j) \underline{P}_Y(1) \\ &= F_X(x) + (1 - F_X(x)) \underline{P}_Y(1). \end{aligned}$$

On the other hand, it holds that

$$\begin{aligned} \underline{P}(r(X,Y,d) \leq x) &= \sum_{x_j \leq x} P_X(x_j) + \min_{i=1,2} \left\{ \sum_{x_j > x} P_X(x_j) P_Y^{p_i}(1) \right\} \\ &= \sum_{x_j \leq x} P_X(x_j) + \sum_{x_j > x} P_X(x_j) \underline{P}_Y(1) \\ &= F_X(x) + (1 - F_X(x)) \underline{P}_Y(1), \end{aligned}$$

therefore, the statement follows. □

Remark 2. In analogy to Remark 1, the lower cumulative distribution function $\underline{F}^{oo}_{r(X,Y,d)}(x)$ can be associated with an upper cumulative distribution function, defined as $\overline{F}^{oo}_{r(X,Y,d)}(x) := 1 - Bel^{oo}(r(X,Y,d) > x)$. Proceeding as in the proof of Theorem 4, it is easy to prove that

$$\overline{F}^{oo}_{r(X,Y,d)}(x) = \overline{F}_{r(X,Y,d)}(x) = \begin{cases} 1 - F_X(x) & \text{if } 0 \leq x < d, \\ 1 - (1 - F_X(x)) \underline{P}_Y(0) & \text{if } x \geq d, \end{cases}$$

where the first branch vanishes if $d = 0$. Hence, we get that the p-boxes $(\underline{F}_{r(X,Y,d)}, \overline{F}_{r(X,Y,d)})$ and $(\underline{F}^{oo}_{r(X,Y,d)}, \overline{F}^{oo}_{r(X,Y,d)})$ coincide though $\underline{P}$ and Bel^{oo} are different.

The following proposition shows that $\underline{F}^{oo}_{r(X,Y,d)}(x)$ can be formally regarded as a cumulative distribution function.

Proposition 4. *The function $\underline{F}^{oo}_{r(X,Y,d)}(x)$ satisfies the following properties:*

(i) *it is non-decreasing;*
(ii) *it is right-continuous;*
(iii) *it is piece-wise constant;*
(iv) $\lim_{x \to +\infty} \underline{F}^{oo}_{r(X,Y,d)}(x) = 1.$

Proof. The proof is an immediate consequence of Theorem 4 and the properties of F_X. □

We finally derive the following corollary, stating that the optimal retention level selection problems under $\underline{P}$ and Bel^{oo} coincide.

Corollary 1. *The following statements hold:*

(i) $\underline{\mathrm{VaR}}^{oo}_{\alpha}(r(X,Y,d)) = \underline{\mathrm{VaR}}_{\alpha}(r(X,Y,d))$;
(ii) $\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)] = \underline{\mathbb{E}}_{\mathcal{P}}[Z(X,Y,d)]$.

Proof. Theorems 3 and 4 imply *(i)* and, so, $Z^{oo}(X,Y,d) = Z(X,Y,d)$. Hence, by (13) and (16) we immediately derive the validity of *(ii)*. □

The following example shows that the random variables $Z^i(X,Y,d)$ and $Z^{oo}(X,Y,d)=Z(X,Y,d)$ are generally different, thus $\mathbb{C}_{Bel^i}[Z^i(X,Y,d)]$ and $\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)]=\underline{\mathbb{E}}_{\mathcal{P}}[Z(X,Y,d)]$ lead to two different optimization problems. In particular, since no containment relation holds between **core**(Bel^i) and $\mathcal{P}$, in general, no dominance relation can be established between $\mathbb{C}_{Bel^i}[Z^i(X,Y,d)]$ and $\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)]=\underline{\mathbb{E}}_{\mathcal{P}}[Z(X,Y,d)]$: this makes Bel^i unsuitable in our reinsurance application.

Example 6. Let X, Y, $\mathcal{P}$, α, θ, r_{coc}, $\pi(X)$ as in Example 5. Denote

$$\mathcal{X}\times\mathcal{Y}=\{\underbrace{(0,1)}_{=a_1},\underbrace{(0,0)}_{=a_2},\underbrace{(100,1)}_{=a_3},\underbrace{(100,0)}_{=a_4},\underbrace{(1000,1)}_{=a_5},\underbrace{(1000,0)}_{=a_6}\},$$

and let $A_{i_1\cdots i_k}=\{a_{i_1},\ldots,a_{i_k}\}$.

The D_2-optimal X-preserving outer DS-approximation Bel^{oo} of $\underline{P}$ has Möbius inverse m^{oo} such that

$$\begin{aligned}
&m^{oo}(A_1)=\tfrac{720}{1000},\ m^{oo}(A_2)=\tfrac{90}{1000},\ m^{oo}(A_{12})=\tfrac{90}{1000},\\
&m^{oo}(A_3)=\tfrac{48}{1000},\ m^{oo}(A_4)=\tfrac{6}{1000},\ m^{oo}(A_{34})=\tfrac{6}{1000},\\
&m^{oo}(A_5)=\tfrac{32}{1000},\ m^{oo}(A_6)=\tfrac{4}{1000},\ m^{oo}(A_{56})=\tfrac{4}{1000},
\end{aligned}$$

and zero elsewhere. Moreover, by Corollary 1 and Example 5 it follows that $d^*=100$ is the optimal retention level.

We point out that switching to the D_2-optimal inner DS-approximation of $\underline{P}$ leads, in general, to a different optimization problem. Indeed, the D_2-optimal inner DS-approximation Bel^i of $\underline{P}$ has Möbius inverse m^i such that

$$\begin{aligned}
&m^{i}(A_1)=\tfrac{725}{1000},\ m^{i}(A_2)=\tfrac{95}{1000},\ m^{i}(A_{12})=\tfrac{80}{1000},\\
&m^{i}(A_3)=\tfrac{51}{1000},\ m^{i}(A_4)=\tfrac{9}{1000},\\
&m^{i}(A_5)=\tfrac{34}{1000},\ m^{i}(A_6)=\tfrac{6}{1000},
\end{aligned}$$

and zero elsewhere. A straightforward computation shows that

$$\mathbb{C}_{Bel^i}[-r(X,Y,d)]=\begin{cases}-0.085d-6.9 & \text{if } 0\le d<100,\\ -0.034d-12 & \text{if } 100\le d<1000,\\ -46 & \text{if } d\ge 1000,\end{cases}$$

therefore, we get

$$\pi_R^i(d)=\begin{cases}-0.0935d+43.01 & \text{if } 0\le d<100,\\ -0.0374d+37.4 & \text{if } 100\le d<1000,\\ 0 & \text{if } d\ge 1000.\end{cases}$$

Moreover, $\underline{F}^i_{r(X,Y,d)}(x)$ has the following definitions, according to the value of d

$d=0$

$$\underline{F}^i_{r(X,Y,d)}(x)=\begin{cases}0 & \text{if } x<0,\\ 0.985 & \text{if } 0\le x<100,\\ 0.994 & \text{if } 100\le x<1000,\\ 1 & \text{if } x\ge 1000,\end{cases}$$

$0<d<100$

$$\underline{F}^i_{r(X,Y,d)}(x)=\begin{cases}0 & \text{if } x<0,\\ 0.9 & \text{if } 0\le x<d,\\ 0.985 & \text{if } d\le x<100,\\ 0.994 & \text{if } 100\le x<1000,\\ 1 & \text{if } x\ge 1000,\end{cases}$$

$d=100$

$$\underline{F}^i_{r(X,Y,d)}(x)=\begin{cases}0 & \text{if } x<0,\\ 0.9 & \text{if } 0\le x<100,\\ 0.994 & \text{if } 100\le x<100,\\ 1 & \text{if } x\ge 1000,\end{cases}$$

$100<d<1000$

$$\underline{F}^i_{r(X,Y,d)}(x)=\begin{cases}0 & \text{if } x<0,\\ 0.9 & \text{if } 0\le x<100,\\ 0.96 & \text{if } 100\le x<d,\\ 0.994 & \text{if } d\le x<1000,\\ 1 & \text{if } x\ge 1000,\end{cases}$$

$d\ge 1000$

$$\underline{F}^i_{r(X,Y,d)}(x)=\begin{cases}0 & \text{if } x<0,\\ 0.9 & \text{if } 0\le x<100,\\ 0.96 & \text{if } 100\le x<1000,\\ 1 & \text{if } x\ge 1000.\end{cases}$$

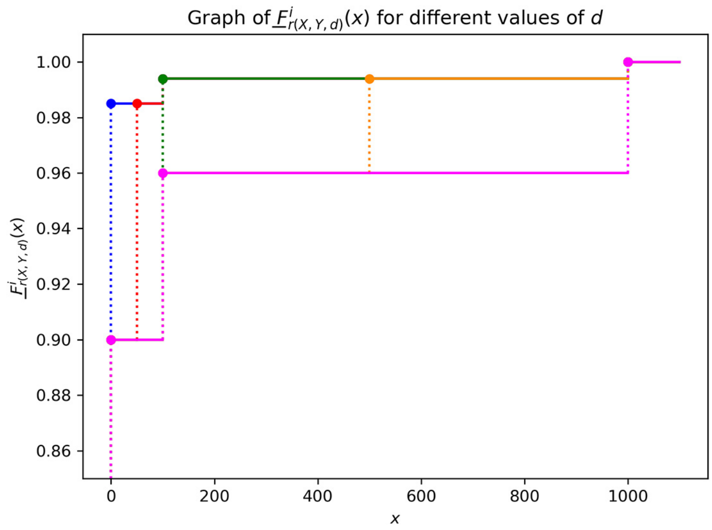

Fig. 2. Graph of $\underline{F}^i_{r(X,Y,d)}(x)$: $d = 0$ in blue; $0 < d < 100$ in red; $d = 100$ in green; $100 < d < 1000$ in orange; $d \geq 1000$ in magenta.

Fig. 2 shows the graph of $\underline{F}^i_{r(X,Y,d)}(x)$ for different values of d.

Though the definitions of $\underline{F}^i_{r(X,Y,d)}(x)$ differ from $\underline{F}^{oo}_{r(X,Y,d)}(x)$, it holds that $\underline{\mathrm{VaR}}^i_\alpha(r(X,Y,d)) = \underline{\mathrm{VaR}}^{oo}_\alpha(r(X,Y,d)) = \underline{\mathrm{VaR}}_\alpha(r(X,Y,d))$, thus we get that

$$\mathbb{C}_{Bel^i}[Z^i(X,Y,d)] = \begin{cases} \left(\frac{0.0935}{0.93} - 0.085\right)d + \left(\frac{0.59}{0.93} - 6.9\right) & \text{if } 0 \leq d < 100, \\ \left(-\frac{0.0326}{0.93} - 0.034\right)d + \left(\frac{13.2}{0.93} - 12\right) & \text{if } 100 \leq d < 1000, \\ -\frac{19.4}{0.93} - 46 & \text{if } d \geq 1000. \end{cases}$$

From Fig. 3 it is easily seen that both $\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)] = \underline{\mathbb{E}}_{\mathcal{P}}[Z(X,Y,d)]$ and $\mathbb{C}_{Bel^i}[Z^i(X,Y,d)]$ have a global maximum at $d^* = 100$.

Nevertheless, the maximum Choquet expected profits computed with respect to Bel^{oo} and Bel^i differ, as $\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d^*)] = \underline{\mathbb{E}}_{\mathcal{P}}[Z(X,Y,d^*)] \approx -4.38$ and $\mathbb{C}_{Bel^i}[Z^i(X,Y,d^*)] \approx -4.71$.

5. Characterization of the optimal retention level d^*

The absence of containment relationships between $\mathcal{P}$ and **core**(Bel^i) impairs the use of Bel^i in our reinsurance application, hence, we will focus on Bel^{oo} only. In this section we present the study of the optimization problem

$$\max_{d \geq 0} \mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)], \tag{17}$$

where $\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)]$ is defined as in (16).

First of all, we have to determine the expression of $\underline{\mathrm{VaR}}^{oo}_\alpha(r(X,Y,d))$ that, according to (15), can be computed through the *generalized inverse* (see [16]) of $\underline{F}^{oo}_{r(X,Y,d)}(x)$, that is

$$\underline{\mathrm{VaR}}^{oo}_\alpha(r(X,Y,d)) = (\underline{F}^{oo}_{r(X,Y,d)})^{-1}(\alpha).$$

Since $\underline{F}^{oo}_{r(X,Y,d)}(x)$ is non-decreasing, right-continuous and piece-wise constant by Proposition 4, its generalized inverse $(\underline{F}^{oo}_{r(X,Y,d)})^{-1}(\alpha)$ is non-decreasing, left-continuous and piece-wise constant, and ranges in $\mathcal{X}$ for $\alpha \in (0,1)$.

We first notice that for $0 < \alpha \leq F_X(0)$, then $\underline{\mathrm{VaR}}^{oo}_\alpha(r(X,Y,d)) = 0$, thus we can limit to consider $\alpha \in (F_X(0),1)$. The following proposition expresses $\underline{\mathrm{VaR}}^{oo}_\alpha(r(X,Y,d))$ as a function of $d \geq 0$, for a fixed value of α.

Proposition 5. *For $\alpha \in (F_X(0),1)$, the following statements hold:*

(i) if $\underline{P}_Y(1) = 0$, then

$$\underline{\mathrm{VaR}}^{oo}_\alpha(r(X,Y,d)) = F_X^{-1}(\alpha);$$

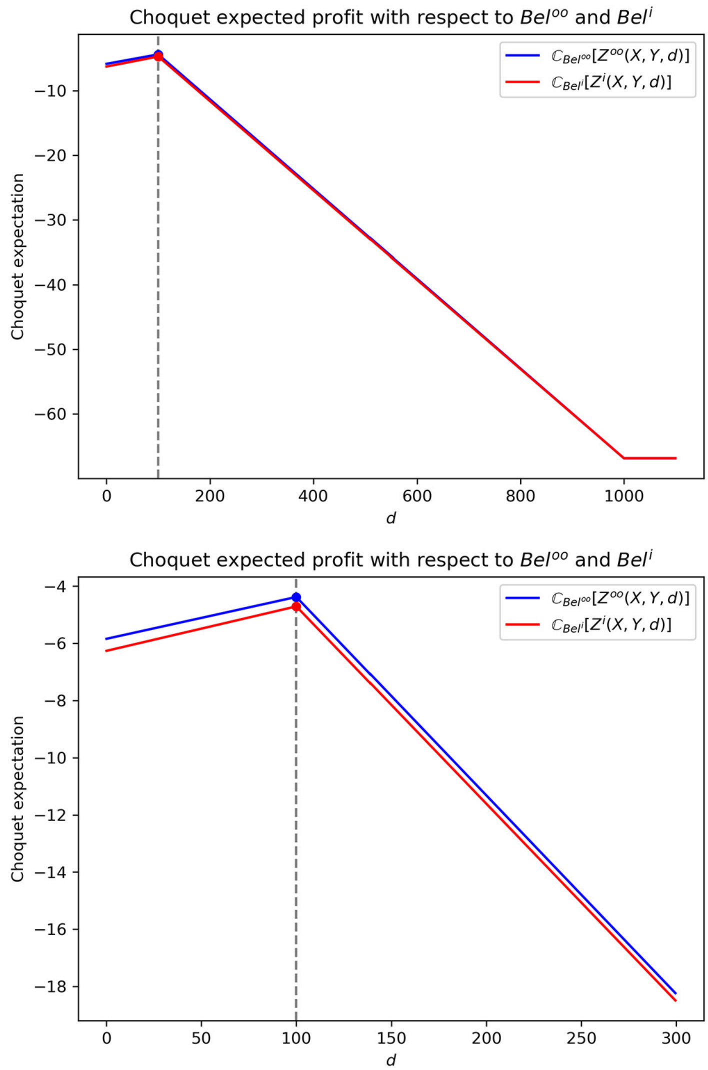

Fig. 3. Graph of $\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)] = \underline{\mathbb{E}}_{\mathcal{P}}[Z(X,Y,d)]$ and $\mathbb{C}_{Bel^{i}}[Z^{i}(X,Y,d)]$ as functions of d (upper plot for $d \in [0,1100]$, lower plot for $d \in [0,300]$).

(ii) if $0 < \underline{P}_Y(1) < \alpha$, *then*

$$\underline{\mathrm{VaR}}^{oo}_{\alpha}(r(X,Y,d)) = \begin{cases} F_X^{-1}\left(\frac{\alpha-\underline{P}_Y(1)}{1-\underline{P}_Y(1)}\right) & \text{if } 0 \leq d < F_X^{-1}\left(\frac{\alpha-\underline{P}_Y(1)}{1-\underline{P}_Y(1)}\right), \\ d & \text{if } F_X^{-1}\left(\frac{\alpha-\underline{P}_Y(1)}{1-\underline{P}_Y(1)}\right) \leq d < F_X^{-1}(\alpha), \\ F_X^{-1}(\alpha) & \text{if } d \geq F_X^{-1}(\alpha); \end{cases}$$

(iii) if $\alpha \leq \underline{P}_Y(1) \leq 1$, *then*

$$\underline{\mathrm{VaR}}^{oo}_{\alpha}(r(X,Y,d)) = \begin{cases} d & \text{if } 0 \leq d < F_X^{-1}(\alpha), \\ F_X^{-1}(\alpha) & \text{if } d \geq F_X^{-1}(\alpha), \end{cases}$$

where $F_X^{-1}(\cdot)$ *denotes the generalized inverse of* $F_X(\cdot)$.

Proof. We first determine $(\underline{F}^{oo}_{r(X,Y,d)}(x))^{-1}(\alpha)$ as a function of α. Set

$$G(x) := F_X(x) + (1 - F_X(x))\underline{P}_Y(1).$$

Since for $x \geq d$ it holds that

$$G(x) \geq \alpha \Longleftrightarrow F_X(x)(1 - \underline{P}_Y(1)) \geq \alpha - \underline{P}_Y(1) \Longleftrightarrow F_X(x) \geq \frac{\alpha - \underline{P}_Y(1)}{(1 - \underline{P}_Y(1))}$$

we obtain

$$(\underline{F}^{oo}_{r(X,Y,d)}(x))^{-1}(\alpha) = \begin{cases} G^{-1}(\alpha) & \text{if } \alpha \in [G(d), 1), \\ d & \text{if } \alpha \in [F_X(d), G(d)), \\ F_X^{-1}(\alpha) & \text{if } \alpha \in (F_X(0), F_X(d)), \end{cases}$$

$$= \begin{cases} F_X^{-1}\left(\frac{\alpha - \underline{P}_Y(1)}{1 - \underline{P}_Y(1)}\right) & \text{if } \alpha \in [G(d), 1), \\ d & \text{if } \alpha \in [F_X(d), G(d)), \\ F_X^{-1}(\alpha) & \text{if } \alpha \in (F_X(0), F_X(d)). \end{cases}$$

Next we express $\underline{\mathrm{VaR}}^{oo}_\alpha(r(X,Y,d)) = (\underline{F}^{oo}_{r(X,Y,d)}(x))^{-1}(\alpha)$ as a function of d, taking $\alpha \in (F_X(0), 1)$ as a fixed constant. We first notice that if $\underline{P}_Y(1) = 0$, then $\underline{F}^{oo}_{r(X,Y,d)} \equiv F_X$, therefore statement *(i)* follows.

Hence, suppose $\underline{P}_Y(1) \in (0, 1]$. In this case, considering the above expression of $(\underline{F}^{oo}_{r(X,Y,d)}(x))^{-1}(\alpha)$, we have that

$$0 < \frac{\alpha - \underline{P}_Y(1)}{1 - \underline{P}_Y(1)} < 1 \Longleftrightarrow \underline{P}_Y(1) < \alpha,$$

and also

$$\alpha > G(d) \Longleftrightarrow d \leq F_X^{-1}\left(\frac{\alpha - \underline{P}_Y(1)}{1 - \underline{P}_Y(1)}\right),$$

and hence statements *(ii)* and *(iii)* follow. □

Now, we are ready to characterize the optimal retention level d^*. Taking into account equations (14), (16) and Theorem 3, for $d < x_t$, we derive that

$$\begin{aligned} &\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)] = \tfrac{\pi(X)}{1-r_{coc}} - \mathbb{E}_{P_X}[X] \\ &- \tfrac{r_{coc}}{1-r_{coc}}\left[\eta(1 - F_X(x_{k(d)}))\mathbb{E}_{P_X}[X - d | X > x_{k(d)}] + \underline{\mathrm{VaR}}^{oo}_\alpha(r(X,Y,d))\right], \end{aligned} \tag{18}$$

where

$$\eta := \left(1 + \frac{\theta}{r_{coc}}\right)\underline{P}_Y(1),$$

and $x_{k(d)}$ is defined as in Theorem 3. Note that η vanishes if and only if $\underline{P}_Y(1) = 0$.

Hence, defining

$$g(d) := \eta(1 - F_X(x_{k(d)}))\mathbb{E}_{P_X}[X - d | X > x_{k(d)}] + \underline{\mathrm{VaR}}^{oo}_\alpha(r(X,Y,d)), \tag{19}$$

we get that the optimization problem (17) is equivalent to

$$\min_{d \geq 0} g(d). \tag{20}$$

Let us notice that, for $d \geq x_t$ we have that $r(X,Y,d) \equiv X$, which implies that $g(d) = F_X^{-1}(\alpha)$. Thus, the optimization problem (20) can actually be solved for $0 \leq d \leq x_t$, and an optimal retention level $d^* = x_t$ can be interpreted as if no reinsurance takes place. In particular, $g(d)$ is easily seen to be a piece-wise linear continuous function on the compact set $[0, x_t]$, which always admits a global minimum by Weierstrass theorem.

We know that, for $0 < \alpha \leq F_X(0)$, we have that $\underline{\mathrm{VaR}}^{oo}_\alpha(r(X,Y,d)) = 0$, and this implies that $g(d)$ is minimized for $d^* = x_t$. Therefore, in what follows we assume that $\alpha \in (F_X(0), 1)$.

Theorem 5. *Let d^* be the retention level maximizing $\mathbb{C}_{Bel^{oo}}[Z^{oo}(X,Y,d)]$, defined as in equation (18). For $\alpha \in (F_X(0), 1)$, let $\underline{k}$ and $\overline{k}$ be the indexes in $\{1, \ldots, t\}$ such that $x_{\underline{k}} = F_X^{-1}\left(\frac{\alpha - \underline{P}_Y(1)}{1 - \underline{P}_Y(1)}\right)$ if $\underline{P}_Y(1) < 1$ and $\underline{k} = 1$ otherwise, while $x_{\overline{k}} = F_X^{-1}(\alpha)$. Then, the following statements hold:*

(i) if $\underline{P}_Y(1) = 0$, then $d^* = x_t$;

(ii) *if* $\underline{P}_Y(1) > 0$, *then*
   (a) $x_{\underline{k}} = x_{\overline{k}}$ *or* $\eta^{-1} < 1 - F_X(x_{\overline{k}-1})$ *implies* $d^* = x_t$;
   (b) $\eta^{-1} \geq 1 - F_X(x_{\overline{k}-1})$ *implies*

$$d^* = \begin{cases} x_{i^*} & \text{if } g(x_{i^*}) \leq x_{\overline{k}}, \\ x_t & \text{otherwise}; \end{cases}$$

with $i^* = \min\{i \in \{\underline{k}, \dots, \overline{k}-1\} : \eta^{-1} \geq 1 - F_X(x_i)\}$ *and* $g(d)$ *is defined as in equation* (19).

Proof. We minimize $g(d)$ over $0 \leq d \leq x_t$. We first notice that if $\underline{P}_Y(1) = 0$, then from equation (19) and statement *(i)* of Proposition 5 we get that

$$g(d) = \underline{\text{VaR}}^{oo}_\alpha(r(X, Y, d)) = F_X^{-1}(\alpha),$$

therefore statement *(i)* follows. Hence, suppose $\underline{P}_Y(1) \in (0, 1]$.

Statement *(ii)*. First assume $0 < \underline{P}_Y(1) < \alpha$. We have that

$$g(d) = \begin{cases} \eta(1 - F_X(x_i))\mathbb{E}_{P_X}[X - d | X > x_i] & \text{if } d \in [x_i, x_{i+1}), i = 1, \dots, t-1, \\ 0 & \text{if } d = x_t, \end{cases}$$
$$+ \begin{cases} x_{\underline{k}} & \text{if } 0 \leq d < x_{\underline{k}}, \\ d & \text{if } x_{\underline{k}} \leq d < x_{\overline{k}}, \\ x_{\overline{k}} & \text{if } x_{\overline{k}} \leq d \leq x_t. \end{cases}$$

If $\underline{k} = \overline{k}$, then $g(d)$ is always decreasing, so $d^* = x_t$. Thus, suppose $\underline{k} < \overline{k}$. In this case, we have that

$$g'(d) = \begin{cases} -\eta(1 - F_X(x_i)) + 1 & \text{if } d \in (x_i, x_{i+1}), i \in \{\underline{k}, \dots, \overline{k}-1\}, \\ -\eta(1 - F_X(x_i)) & \text{if } d \in (x_i, x_{i+1}), i \in \{1, \dots, t-1\} \setminus \{\underline{k}, \dots, \overline{k}-1\}, \end{cases}$$

so, $g(d)$ can be increasing only for $d \in (x_i, x_{i+1})$, $i \in \{\underline{k}, \dots, \overline{k}-1\}$. If $\eta^{-1} < 1 - F_X(x_{\overline{k}-1})$ then $g(d)$ is always decreasing, so $d^* = x_t$. Thus, suppose there is at least an index i such that $\eta^{-1} \geq 1 - F_X(x_i)$ and let i^* be the minimum index in $\{\underline{k}, \dots, \overline{k}-1\}$ with $\eta^{-1} \geq 1 - F_X(x_{i^*})$. Notice that, for $i = i^* + 1, \dots, \overline{k}-1$, it holds that

$$-\eta(1 - F_X(x_{i^*})) + 1 < \eta(1 - F_X(x_i)) + 1,$$

thus $d^* = x_{i^*}$ if $g(x_{i^*}) \leq x_{\overline{k}}$ and $d^* = x_t$ otherwise. Hence, conditions *(a)* and *(b)* follow.

Now assume $\alpha \leq \underline{P}_Y(1) \leq 1$. We have that

$$g(d) = \begin{cases} \eta(1 - F_X(x_i))\mathbb{E}_{P_X}[X - d | X > x_i] & \text{if } d \in [x_i, x_{i+1}), i = 1, \dots, t-1, \\ 0 & \text{if } d = x_t, \end{cases}$$
$$+ \begin{cases} d & \text{if } 0 \leq d \leq x_{\overline{k}}, \\ x_{\overline{k}} & \text{if } x_{\overline{k}} \leq d \leq x_t. \end{cases}$$

Analogously to the previous step, we have that

$$g'(d) = \begin{cases} -\eta(1 - F_X(x_i)) + 1 & \text{if } d \in (x_i, x_{i+1}), i \in \{1, \dots, \overline{k}-1\}, \\ -\eta(1 - F_X(x_i)) & \text{if } d \in (x_i, x_{i+1}), i \in \{1, \dots, t-1\} \setminus \{1, \dots, \overline{k}-1\}, \end{cases}$$

so, $g(d)$ can be increasing only for $d \in (x_i, x_{i+1})$, $i \in \{1, \dots, \overline{k}-1\}$. If $\eta^{-1} < 1 - F_X(x_{\overline{k}-1})$ then $g(d)$ is always decreasing, so $d^* = x_t$. Thus, suppose there is at least an index i such that $\eta^{-1} \geq 1 - F_X(x_i)$ and let i^* be the minimum index in $\{1, \dots, \overline{k}-1\}$ with $\eta^{-1} \geq 1 - F_X(x_{i^*})$. Notice that, for $i = i^* + 1, \dots, \overline{k}-1$, it holds that

$$-\eta(1 - F_X(x_{i^*})) + 1 < -\eta(1 - F_X(x_i)) + 1,$$

thus $d^* = x_{i^*}$ if $g(x_{i^*}) \leq x_{\overline{k}}$ and $d^* = x_t$ otherwise. Again conditions *(a)* and *(b)* follow. □

The following example shows an application of previous theorem.

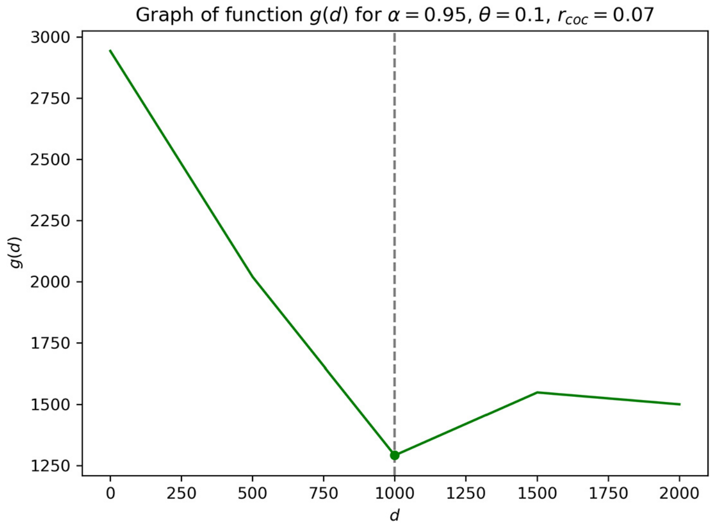

Fig. 4. Graph of function $g(d)$ for $\alpha = 0.95$, $\theta = 0.1$, $r_{coc} = 0.07$.

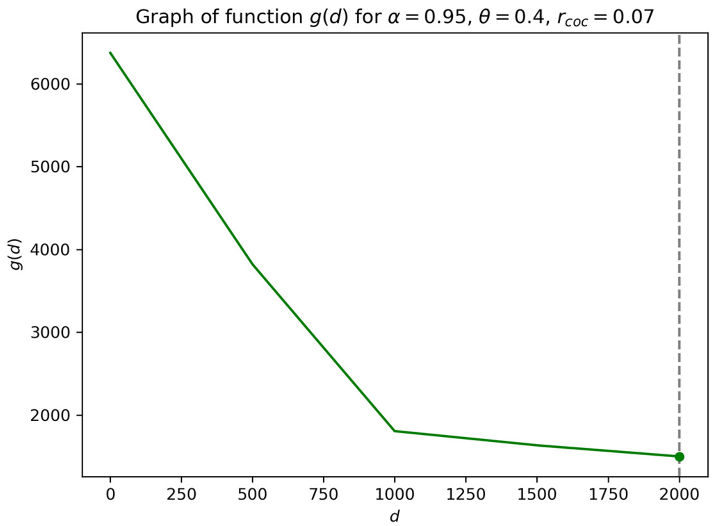

Fig. 5. Graph of function $g(d)$ for $\alpha = 0.95$, $\theta = 0.4$, $r_{coc} = 0.07$.

Example 7. Consider X with range $\mathcal{X} = \{0, 500, 1000, 1500, 2000\}$ and take the probability distributions such that $P_X(0) = P_X(2000) = \frac{5}{100}$, $P_X(500) = P_X(1500) = \frac{20}{100}$, $P_X(1000) = \frac{50}{100}$, and $P_Y^p(1) = p$, $P_Y^p(0) = 1 - p$, with $p \in \left[\frac{8}{10}, \frac{9}{10}\right]$, therefore $\underline{P}_Y(1) = \frac{8}{10}$.

Taking the parameters $\alpha = 0.95$, $\theta = 0.1$, $r_{coc} = 0.07$, we have that $\eta^{-1} = 0.5147$, $x_{\underline{k}} = 1000$ and $x_{\overline{k}} = 1500$. Therefore, since $\eta^{-1} \geq 0.25 = 1 - F_X(1000)$, we get that $x_{i^*} = 1000$, moreover, since $g(x_{i^*}) = 1291.4286 \leq x_{\overline{k}}$, we conclude that $d^* = 1000$, in agreement with condition *(ii.b)* of Theorem 5. Fig. 4 shows the function $g(d)$ and the optimal retention level d^*.

On the other hand, if we take the parameters $\alpha = 0.95$, $\theta = 0.4$, $r_{coc} = 0.07$, we have that $\eta^{-1} = 0.1862$, $x_{\underline{k}} = 1000$ and $x_{\overline{k}} = 1500$. Therefore, since $\eta^{-1} < 0.25 = 1 - F_X(1000)$, we conclude that $d^* = 2000$, in agreement with condition *(ii.a)* of Theorem 5. Fig. 5 shows the function $g(d)$ and the optimal retention level d^*.

Hence, in this example, we see that the optimal retention level is $d^* = 1000$ when the safety loading is $\theta = 0.1$, while it becomes $d^* = 2000$ when θ increases to 0.4, indicating that no reinsurance takes place.

6. Conclusion

This paper introduces ambiguity in the randomized reinsurance stop-loss treaty formalized in [2], limiting to a discrete aggregate loss X. Following [2] we require independence between X and Y under P_X and any element of $\mathcal{P}_Y$. The independence assumption is easy to justify when Y is the indicator of an exogenous event, while the justification is weak when Y is the indicator of a default event.

Given the set $\mathcal{P}$ of joint distributions of (X, Y), we showed that switching to the D_2-optimal X-preserving outer approximation Bel^{oo} of $\underline{P} = \min \mathcal{P}$, allows us to relax the independence assumption on $\mathcal{P}$ and yet does not introduce spurious information in the retention level selection problem. In this concern, Theorem 4 and Corollary 1 are of particular relevance, since they show that the independence assumption in the original set $\mathcal{P}$ of joint distributions is not too strong as, relaxing it, we get the same optimization problem.

The use of Bel^{oo} in computing Choquet expectations and the ambiguous value-at-risk $\underline{\text{VaR}}_\alpha^{oo}(r(X, Y, d))$ implement a pessimistic attitude towards ambiguity, that generally leads to a cautious estimate of the optimal retention level d^*, in comparison to the precise case. As highlighted in Remarks 1 and 2, switching to upper expectations and upper cumulative distribution functions, an optimistic attitude can be modeled, though the duality with respect to the pessimistic case does not hold, in general.

We point out that $\mathbf{core}(Bel^{oo})$ naturally leads to consider a *coherent risk measure* in the sense of [3], in place of the ambiguous value-at-risk. In turn, coherent risk measures have been thoroughly analyzed in the setting of imprecise probabilities (see, e.g., [30,40]), so, their use in the present model seems to be an interesting line of future research.

The present model can be extended in two directions, which are meaningful from a reinsurance point of view: *(i)* by considering a random variable Y ranging in $\mathcal{Y} = \{0, \ldots, n\}$, endowed with a closed and convex set of probability distributions $\mathcal{P}_Y$; *(ii)* by taking a random variable X ranging in $\mathcal{X} = [0, +\infty)$, endowed with a continuous probability distribution. In the case *(i)*, Y expresses one of $n + 1$ reinsurance scenarios that determine the computation of the retained loss $r(X, Y, d)$: an analog of Proposition 1 can be proved, so we still have a closed and convex set of joint probability distributions $\mathcal{P}$, under independence of X and Y. In this case, the issue of computing an inner [X-preserving outer] DS-approximation is an open problem. On the other hand, the case *(ii)* can be faced by introducing suitable distortion models that give rise to belief functions and by relying on the framework of finitely additive probabilities, in the spirit of [23,24].

Still referring to finite-range X and Y, a more involved situation is met if both X and Y are endowed with closed and convex sets of probability distributions $\mathcal{P}_X$ and $\mathcal{P}_Y$, respectively. Indeed, even by referring to binary X and Y, it is easy to show that the set $\mathcal{P}$ of independent products generated by $\mathcal{P}_X$ and $\mathcal{P}_Y$ may be not convex. In particular, this paves the way to the several definitions of independence arising in the context of imprecise probabilities [10,38,41].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

The research has been partially supported by PNRR MUR project `PE0000013-FAIR`. The first and third authors are members of the GNAMPA-INdAM research group. The last two authors acknowledge financial support from Sapienza University of Rome project `RP120172B81E33DD`.

References

[1] H. Albrecher, J. Beirlant, J.L. Teugels, Reinsurance: Actuarial and Statistical Aspects, Wiley Series in Probability and Statistics, Wiley, 2017.
[2] H. Albrecher, A. Cani, On randomized reinsurance contracts, Insur. Math. Econ. 84 (2019) 67–78.
[3] P. Artzner, F. Delbaen, J.-M. Eberc, D. Heath, Coherent measures of risk, Math. Finance 9 (3) (1999) 203–228.
[4] P. Baroni, P. Vicig, An uncertainty interchange format with imprecise probabilities, Int. J. Approx. Reason. 40 (3) (2005) 147–180.
[5] G.W. Brier, Verification of forecasts expressed in terms of probability, Mon. Weather Rev. 78 (1) (1950) 1–3.
[6] Y. Censor, S.A. Zenios, Parallel Optimization: Theory, Algorithms and Applications, Oxford University Press, 1997.
[7] A. Chateauneuf, J.-Y. Jaffray, Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, Math. Soc. Sci. 17 (3) (1989) 263–283.
[8] A. Cinfrignini, D. Petturiti, B. Vantaggi, Envelopes of equivalent martingale measures and a generalized no-arbitrage principle in a finite setting, Ann. Oper. Res. 321 (1) (2023) 103–137.
[9] I. Couso, S. Moral, Independence concepts in evidence theory, Int. J. Approx. Reason. 51 (7) (2010) 748–758.
[10] I. Couso, S. Moral, P. Walley, A survey of concepts of independence for imprecise probabilities, Risk Decis. Policy 5 (2) (2000) 165–181.
[11] F.G. Cozman, Constructing sets of probability measures through Kuznetsov's independence condition, in: Proceedings of ISIPTA'2001, 2001, pp. 104–111.
[12] G. de Cooman, E. Miranda, M. Zaffalon, Independent natural extension, Artif. Intell. 175 (12) (2011) 1911–1950.
[13] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (2) (1967) 325–339.
[14] D. Denneberg, Non-Additive Measure and Integral, Kluwer Academic Publishers, 1994.
[15] T. Denœux, Constructing belief functions from sample data using multinomial confidence regions, Int. J. Approx. Reason. 42 (3) (2006) 228–252.
[16] P. Embrechts, M. Hofert, A note on generalized inverses, Math. Methods Oper. Res. 77 (3) (2013) 423–432.
[17] S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, Constructing probability boxes and Dempster-Shafer structures, Technical Report SAND2002-4015, Sandia National Laboratories, 2003.

[18] M. Grabisch, Set Functions, Games and Capacities in Decision Making, Springer International Publishing, 2016.
[19] R. Kaas, M. Goovaerts, J. Dhaene, M. Denuit, Modern Actuarial Risk Theory: Using R, 2nd edition, Springer, Berlin, Heidelberg, 2008.
[20] I. Levi, The Enterprise of Knowledge, MIT Press, Cambridge, 1980.
[21] E. Miranda, I. Montes, A. Presa, Inner approximations of credal sets by non-additive measures, in: D. Ciucci, I. Couso, J. Medina, D. Slezak, D. Petturiti, B. Bouchon-Meunier, R.R. Yager (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems, Springer International Publishing, Cham, 2022, pp. 743–756.
[22] E. Miranda, I. Montes, P. Vicig, On the selection of an optimal outer approximation of a coherent lower probability, Fuzzy Sets Syst. 424 (2021) 1–36.
[23] I. Montes, E. Miranda, S. Destercke, Unifying neighbourhood and distortion models: part I – new results on old models, Int. J. Gen. Syst. 49 (6) (2020) 602–635.
[24] I. Montes, E. Miranda, S. Destercke, Unifying neighbourhood and distortion models: part II – new models and synthesis, Int. J. Gen. Syst. 49 (6) (2020) 636–674.
[25] I. Montes, E. Miranda, P. Vicig, 2-monotone outer approximations of coherent lower probabilities, Int. J. Approx. Reason. 101 (2018) 181–205.
[26] I. Montes, E. Miranda, P. Vicig, Outer approximating coherent lower probabilities with belief functions, Int. J. Approx. Reason. 110 (2019) 1–30.
[27] S. Moral, Discounting imprecise probabilities, in: E. Gil, E. Gil, J. Gil, M.A. Gil (Eds.), The Mathematics of the Uncertain: A Tribute to Pedro Gil, Springer International Publishing, Cham, 2018, pp. 685–697.
[28] S. Moral, A. Cano, M. Gómez-Olmedo, Divergence in Bayesian networks, Entropy 23 (9) (2021) 1122.
[29] R.B. Nelsen, An Introduction to Copulas, Springer Series in Statistics, Springer, New York, NY, 2006.
[30] R. Pelessoni, P. Vicig, Imprecise previsions for risk measurement, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 11 (04) (2003) 393–412.
[31] D. Petturiti, G. Stabile, B. Vantaggi, Addressing ambiguity in randomized reinsurance contracts using belief functions, in: S. Le Hégarat-Mascle, I. Bloch, E. Aldea (Eds.), Belief Functions: Theory and Applications, Springer International Publishing, Cham, 2022, pp. 286–296.
[32] D. Petturiti, B. Vantaggi, How to assess coherent beliefs: a comparison of different notions of coherence in Dempster-Shafer theory of evidence, in: T. Augustin, F. Gagliardi Cozman, G. Wheeler (Eds.), Reflections on the Foundations of Probability and Statistics: Essays in Honor of Teddy Seidenfeld, in: Theory and Decision Library A, vol. 54, Springer International Publishing, Cham, 2022, pp. 161–185.
[33] D. Petturiti, B. Vantaggi, Probability envelopes and their Dempster-Shafer approximations in statistical matching, Int. J. Approx. Reason. 150 (2022) 199–222.
[34] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[35] P.P. Shenoy, An expectation operator for belief functions in the Dempster–Shafer theory, Int. J. Gen. Syst. 49 (1) (2020) 112–141.
[36] M. Troffaes, S. Destercke, Probability boxes on totally preordered spaces for multivariate modelling, Int. J. Approx. Reason. 52 (6) (2011) 767–791.
[37] M.C.M. Troffaes, G. de Cooman, Lower Previsions, Wiley, 2014.
[38] B. Vantaggi, Conditional independence structures and graphical models, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 11 (05) (2003) 545–571.
[39] P. Vicig, Epistemic independence for imprecise probabilities, Int. J. Approx. Reason. 24 (2) (2000) 235–250.
[40] P. Vicig, Financial risk measurement with imprecise probabilities, Int. J. Approx. Reason. 49 (1) (2008) 159–174.
[41] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, 1991.